

This page Is Inserted by IFW Operations
And is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of
The original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
Please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM_protein - nucleic search, using frame_p2n model

Run on: November 25, 2003, 10:20:54 ; Search time 93 Seconds

(without alignments)
2444.218 Million cell updates/sec

Title: US-09-869-433-2
Perfect score: 2630
Sequence: 1 MTTBEXKPFKLSFLWPIH.....AQSALKEQVAVQEDSAPASS 515

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 segs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seg length: 0
Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xjh
-Q=/cgn2_1/USPTO.spool/US09869433/runat_24112003_171459_18152/app_query.fasta-1.711
-DB=Issued_Patents_NA -OPM=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPTC=0
-LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blosome2 -TRANS=human40.cdt
-LIST=45 -DOCLIG=200 -THR SCORE=pct -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=p2n -NDBM=ext HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09869433 @CGN 1.1 56 @runat_24112003_171459_18152 -NCPU=6 -ICPU=3
-NO MMAP -LARBQUDRY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCtUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|---------|------------------------|-------------------|
| 1 | 2617 | 99.5 | 1230025 | 4 US-09-198-452A-1 | Sequence 1, Appl1 |
| C 2 | 1025.5 | 39.0 | 1230025 | 4 US-09-198-452A-1 | Sequence 1, Appl1 |
| 3 | 242 | 9.2 | 280 | 4 US-09-313-294A-5369 | Sequence 5369, Ap |
| 4 | 168 | 6.4 | 265 | 4 US-09-313-294A-3344 | Sequence 3344, Ap |
| C 5 | 135 | 5.1 | 1365 | 4 US-09-252-991A-15706 | Sequence 15706, A |
| 6 | 135 | 5.1 | 1638 | 4 US-09-252-991A-15648 | Sequence 15648, A |
| 7 | 135 | 5.1 | 2859 | 4 US-09-252-991A-15612 | Sequence 15612, A |
| C 8 | 133 | 5.1 | 1830121 | 4 US-09-557-884-1 | Sequence 1, Appl1 |
| 9 | 133 | 5.1 | 1830121 | 4 US-09-643-990A-1 | Sequence 1, Appl1 |
| 10 | 129 | 4.9 | 1467 | 4 US-09-252-991A-1331 | Sequence 1331, Ap |
| 11 | 122.5 | 4.7 | 68750 | 3 US-09-335-409-1 | Sequence 1, Appl1 |
| 12 | 122.5 | 4.7 | 68750 | 4 US-09-568-102-1 | Sequence 1, Appl1 |

| | | | | | |
|------|-------|-----|---------|------------------------|--------------------|
| 13 | 122.5 | 4.7 | 68750 | 4 US-09-567-969-1 | Sequence 1, Appl1 |
| 14 | 122.5 | 4.7 | 68750 | 4 US-09-568-480-1 | Sequence 1, Appl1 |
| 15 | 122.5 | 4.7 | 68750 | 4 US-09-568-486-1 | Sequence 1, Appl1 |
| 16 | 122.5 | 4.7 | 68750 | 4 US-09-568-472-1 | Sequence 1, Appl1 |
| 17 | 122.5 | 4.7 | 68750 | 4 US-09-567-899-1 | Sequence 1, Appl1 |
| 18 | 120.5 | 4.6 | 1605 | 4 US-09-252-991A-14154 | Sequence 14154, A |
| C 19 | 120.5 | 4.6 | 1638 | 4 US-09-252-991A-16400 | Sequence 16400, A |
| 20 | 120.5 | 4.6 | 1824 | 4 US-09-252-991A-14743 | Sequence 14743, A |
| 21 | 120.5 | 4.6 | 2228 | 4 US-09-252-991A-16088 | Sequence 16088, A |
| 22 | 120.5 | 4.6 | 2489 | 4 US-09-252-991A-16193 | Sequence 16193, A |
| 23 | 119.5 | 4.5 | 3403 | 4 US-09-134-001C-2818 | Sequence 2818, Ap |
| 24 | 119 | 4.5 | 2823 | 4 US-09-328-352-4039 | Sequence 4039, Ap |
| C 25 | 118 | 4.5 | 15363 | 4 US-08-961-527-139 | Sequence 139, Ap |
| 26 | 117.5 | 4.5 | 77536 | 4 US-09-410-551B-1 | Sequence 1, Appl1 |
| C 27 | 116 | 4.4 | 1436 | 4 US-08-961-527-296 | Sequence 296, Ap |
| 28 | 115.5 | 4.4 | 1227 | 4 US-09-107-532A-2909 | Sequence 2909, Ap |
| 29 | 115.5 | 4.4 | 1641 | 4 US-09-252-991A-533 | Sequence 533, Ap |
| 30 | 114.5 | 4.4 | 4403765 | 3 US-09-103-840A-2 | Sequence 2, Appl1 |
| 31 | 114.5 | 4.4 | 4411529 | 3 US-09-103-840A-1 | Sequence 1, Appl1 |
| 32 | 114 | 4.3 | 1245 | 4 US-09-328-352-2835 | Sequence 2835, Ap |
| 33 | 113.5 | 4.3 | 1647 | 4 US-09-328-352-2479 | Sequence 2479, Ap |
| C 34 | 113.5 | 4.3 | 2431 | 4 US-09-221-017B-351 | Sequence 351, Ap |
| C 35 | 113.5 | 4.3 | 6755 | 3 US-08-931-999-4 | Sequence 4, Appl1 |
| 36 | 113.5 | 4.3 | 1830121 | 4 US-09-557-884-1 | Sequence 1, Appl1 |
| 37 | 113.5 | 4.3 | 1830121 | 4 US-09-643-990A-1 | Sequence 1800, Ap |
| 38 | 113 | 4.3 | 1437 | 4 US-09-134-001C-1800 | Sequence 1868, Ap |
| 39 | 113 | 4.3 | 1500 | 4 US-09-252-991A-1868 | Sequence 1717, Ap |
| 40 | 113 | 4.3 | 1737 | 4 US-09-252-991A-1717 | Sequence 28, Appl1 |
| 41 | 113 | 4.3 | 3848 | 3 US-09-112-096-28 | Sequence 14, Appl1 |
| 42 | 113 | 4.3 | 5668 | 3 US-09-112-096-14 | Sequence 1, Appl1 |
| C 43 | 113 | 4.3 | 4411529 | 3 US-09-103-840A-1 | Sequence 848, Ap |
| 44 | 112.5 | 4.3 | 1260 | 4 US-09-328-352-848 | Sequence 2007, Ap |
| C 45 | 112 | 4.3 | 1767 | 4 US-09-252-991A-2007 | |

ALIGNMENTS

RESULT 1
US-09-198-452A-1
Sequence 1, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OR INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 1
LENGTH: 1230025
TYPE: DNA
ORGANISM: Chlamydia pneumoniae
FEATURES:
NAME/KEY: misc feature
LOCATION: (1)..(15000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (15001)..(30000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (30001)..(45000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (45001)..(60000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (60001)..(75000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (75001)..(90000)
OTHER INFORMATION: n=a or c or g or t

CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 1
LENGTH: 1230025
TYPE: DNA
ORGANISM: Chlamydia pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(15000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (15001)..(30000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (30001)..(45000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (45001)..(60000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (60001)..(75000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (75001)..(90000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (90001)..(105000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (105001)..(120000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (120001)..(135000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (135001)..(150000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (150001)..(165000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (165001)..(180000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (180001)..(195000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (195001)..(210000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (210001)..(225000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (225001)..(240000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (240001)..(255000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (255001)..(270000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (270001)..(285000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (285001)..(300000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (300001)..(315000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (315001)..(330000)
OTHER INFORMATION: n=a or c or g or t

NAME/KEY: misc feature
LOCATION: (330001)..(345000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (345001)..(360000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (360001)..(375000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (375001)..(390000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (390001)..(405000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (405001)..(420000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (420001)..(435000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (435001)..(450000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (450001)..(465000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (465001)..(480000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (480001)..(495000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (495001)..(510000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (510001)..(525000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (525001)..(540000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (540001)..(555000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (555001)..(570000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (570001)..(585000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (585001)..(600000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (600001)..(615000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (615001)..(630000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (630001)..(645000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (645001)..(660000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (660001)..(675000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (675001)..(690000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature

| |
|---------------------------------------|
| LOCATION: (1690001) .. (705000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |
| LOCATION: (705001) .. (720000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |
| LOCATION: (720001) .. (735000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |
| LOCATION: (735001) .. (750000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |
| LOCATION: (750001) .. (765000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |
| LOCATION: (765001) .. (780000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |
| LOCATION: (780001) .. (795000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |
| LOCATION: (795001) .. (810000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |
| LOCATION: (810001) .. (825000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |
| LOCATION: (825001) .. (840000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |
| LOCATION: (840001) .. (855000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |
| LOCATION: (855001) .. (870000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |
| LOCATION: (870001) .. (885000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |
| LOCATION: (885001) .. (900000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |
| LOCATION: (900001) .. (915000) |
| OTHER INFORMATION: n=a or c or g or t |
| NAME/KEY: misc. feature |

[illegible]

| | | | |
|----|--------|--|--------|
| Oy | 1 | NecthrlystThcGlglulnlyPheProheglyLysLeuAksSerpheLeutrPrpCoilehis | 20 |
| Db | 717505 | ATCGAGTCATCAAGAAGTGAACCCCTTTTCAAGGCTCGGGCATATCTTGCTATTAT | 717446 |
| Oy | 21 | ThrHsgluleuLylsValleuPrometPheLeumecPhephCysIlethrPheasn | 40 |
| Db | 717445 | AAATCAGATTTTCTTAAGTTGTGCATATTTCTACGCCGTTTTTCGTGGCTTTAAC | 717386 |
| Oy | 41 | TyrThrValleuAkpaspThrLysappThrIleuIleValGIyLaIprOgIysErGIyAla | 60 |
| Db | 717385 | TACTGCCGTGTAAAAACAATGAAAGTACTCTGGTATTGTCGGTTCAGATGCTGGGGCA | 717326 |
| Oy | 61 | GluAlaIleProPheIleLysPheTrpLeuValIprCyAlaIleIlePheMetLeu | 80 |
| Db | 717325 | GAAAGCATTCCTTCTTAAGGTTTGGGAATATGTCGCCGAGCTGTATTATTACATATG | 717266 |
| Oy | 81 | IleYrAlaLylsLeuSerAsnIleLeuSerLysGlnAlaleuAubHeYrAlaValGIyThr | 100 |

Db 717265 GCTCATGGGTGGTTATAGGCAAGTGGATCTCTCGGAGATACCGGTTTTTTATTTATGCTTCATGGCC 717206
QY 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
Db 717205 GCATTCCTTGGTTTTTTCTCTCGTTTGGTGATCATTTATCTGTAGGGATACCTTG 717146
QY 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140
Db 717145 CATCTCAACTCTCTCGCTGATTAATTTACAAAGACCTCTTCTTCACAGACCTTCGTGGTTTT 717086
QY 141 ValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTyrValLeuAlaGluLeuTrpGly 160
Db 717085 ATTTGGAATGCTCGTTACTCGAGGTACAGATATTTATACGTATGTCACAGAGCTGTGAGT 717026
QY 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleHisGlu 180
Db 717025 TCGGTGTTCTTTTCGATGTTGTTCTCGGGACATAGCCAACTACAGATTACTCAATTACTGAA 716966
QY 181 AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
Db 716965 GCGGGCCGTTTTTATGCTCTTTCATCATCAGGATTTAAATCTCTCCATATATCGGAGGA 716900
QY 201 ArgAlaIleValTrpAlaSerLysLeuArg---AlaSerValSerGluGlyAlaAspPro 219
Db 716905 GAAATCTCTTATGGATGGGGAAACAAACATTGTGTTGCTTACTCTTCGATCGATGATCC 716846
QY 220 Trp---GlyIleSerLeuArgLeuMetAlaMetThrIleValSerGlyLeuValLeu 238
Db 716845 TGGCACTCTGTAAATCTCAACTTACACATCGATGACACT---TGTTCTGGTTTAATTATG 716788
QY 239 MetAlaSerTyrTrpTrpIleAsnLysAsnVal---LeuThrAspProArgPheTyr 256
Db 716788 ATC-----TGGCTATATAGCGGAGATTCATTCATTGACTATGATACCTTGATC 716740
QY 257 AsnProGluGluMetGlnLysGlyLysGlyAla----- 268
Db 716740 CCTCATCTAGACGCTGCTTGGCAAAAGAGGAGAGCACTACTGCTAATCTAAAGAAAG 716681
QY 269 ---LysProLysMetAsnMetLysAspSerPheLeuTyrLeuAspArgSerProTyrIle 287
Db 716680 AAAAAAAGCTTAAAGCCAAAGCTAGAAACCTTTCTTACACCTCATTCAGTCTGTTATTTA 716621
QY 288 LeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCysIleAsnLeuIleGluValThr 307
Db 716620 TTAGGGCTCGCAATTATGTTCTTACTCATTAATTTGGTATCATCATTAATCGAAGTCGT 716565
QY 308 TrpLysSerGlnLeuLysLeuGlnTyrProAsnMetAsnAspTyrSerGluPheMetGly 327
Db 716560 TGGAAAGATCAAGATTAGCCAGATTTACAGTTCTCCAGTGAATTCATAGGTATAGAGT 716501
QY 328 AsnPheSerPheTrpThrGlyValAlaSerValIleLeuIleMetLeuPheValGlyIleAsn 347
Db 716500 AGAATCACTACCCCTCATATGGCCGCTGTTCTGTATTAAGCAGCTGTACTCTTACCGAGCAG 716441
QY 348 ValIleArgLysPheGlyTrpLeuThrGlyAlaLeuValThrProValMetValLeuLeu 367
Db 716440 TGTATCCGTAATATGGGGATGAGCTGTGGTCTTTTAAGTCACTCCCATTTGGTAATGTTAGT 716381
QY 368 ThrGlyIleValPhePheAlaLeuValIle---PheArgAsnGlnAlaSerGlyLeuVal 386
Db 716380 TCAGACATCGCTCTTTTTCGGAACATATTTTGTGCAAAAAGACACATCTCATATTTTGGG 716321
QY 387 AlaMetPheGlyThrThrProLeuMetLeuAlaValAlaGlyAlaIleGlnAsnIle 406
Db 716320 GGAAGTCTTGGAAATACACCTTGCGCTCTTACGCTCGGACCTGGAGGAGATGCAAAATGTC 716261
QY 407 LeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLysGluMetAlaTyrIlePro 426
Db 716260 CTATCCCGGGGAGCAAAATTTACGTTCTTTGATCAACAAGAAATATGGCTTTATCCCA 716201
QY 427 LeuAspGlnGluGlnLysValLysGlyLysAlaAlaIleAspValValAlaAlaArgPhe 446

| | | | |
|----|--------|--|--------|
| Dd | 716440 | TGTAACCGTAAATGGGAGTAGCCTCGGTGCTTTAAGTCACTCCATTGGGAATCTTAGTT | |
| Qy | 368 | ThrGlyIleValPhePheAlaLeuValIle--PheArgAsnGlnAlaSerGlyLeuVal | 386 |
| Dd | 716380 | TCAGACACGCTCTTTTCGGAACACTATTTTTGCCTGCAAAAAGACACACTCTATATTTTGCG | 716321 |
| Qy | 387 | AlaMetPheGlyThrThrProLeuMetLeuAlaValValGlyAlaIleGlnAsnIle | 406 |
| Dd | 716320 | GGAGTCTTGGAATGACACTCTGTGCTCTAGCTGCTGCTGACCTGGAGGAGATGCCAAATGTC | 716261 |
| Qy | 407 | LeuSerLysSerThrIleTyrlalaLeuPheAspSerThrIleGluMetAlaTyrlIlePro | 426 |
| Dd | 716260 | CATATCCGGGGGACGAATAATTACGTTCTTTGATCAACAAGAATAATGCTTATCCCA | 716201 |
| Qy | 427 | LeuAspGlnGluGlnLysValIleYsgLYsAlaAlaIleAspValValAlaAlaArgPhe | 446 |

```
Db 716200 CTTTCTCCAGAGATATAAATCATGGAAGCCGATTTGATGCTGCTTCAAGATA 716141
Qy 447 G|L|Y|S|S|E|R|G|Y|G|L|A|L|E|U|I|E|G|I|N|G|I|L|E|U|L|E|U|A|I|L|E|C|Y|S|G|I|S|E|R|I|E|G|Y 466
Db 716140 GGAAGATCGAGGCTCTTTATTTATTAACAGAGGCTGTTGATTTCTCTCTGTTGCA 716081
Qy 467 A|A|A|E|T|H|R|P|R|O|Y|L|E|U|A|I|V|A|I|L|E|U|E|U|P|H|E|I|L|E|A|I|L|E|T|R|I|P|L|E|U|A|S|E|R 486
Db 716080 GCAAGTTTAAACGTATGCTGCTTCTTCTCATTTATTAAGTGTGATTCGGCTT 716021
Qy 487 A|A|T|R|I|L|Y|S|E|U|S|L|E|U|P|H|E|U|A|I|A|G|I|N|S|E|R|A|-----L|E|U|L|S 501
Db 716020 GTTGCTATATCGTAAGAAATACACTAGAGCTGATGCTGTAACAACCTTGAA 715961
Qy 502 -----G|U|G|I|N|G|U|A|I|A|G|I|N|U 508
Db 715960 CAACCTAAAGAACCTTCTCTTCAATCGTAGTGAGCCAGAACTGTAGAACAGAA 715901
Qy 509 A|E|P|S|E|R|A| 511
Db 715900 GAAATGCT 715892

RESULT 3
US-09-313-294A-5369
; Sequence 5369, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 5369
; LENGTH: 280
; TYPE: DNA
; ORGANISM: Zea mays
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700350037H1
; NAME/KEY: unsure
; LOCATION: 2
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-5369

Alignment Scores:
Pred. No.: 4.55e-20 Length: 280
Score: 242.00 Matches: 50
Percent Similarity: 77.33% Conservative: 8
Best Local Similarity: 66.67% Mismatches: 17
Query Match: 9.20% Indels: 0
Gaps: 0

US-09-869-433-2 (1-515) x US-09-313-294A-5369 (1-280)
Qy 402 A|A|A|I|E|G|I|N|A|E|U|I|L|E|U|S|E|R|H|Y|S|Y|R|A|L|E|U|P|H|E|A|S|E|R|H|Y|S|G|I|N 421
Db 3 GCAATGCAAGAACATTTTCAGTAAGATGCAAAATACAGTCTGTTGATCCTTGCAAGAG 62
Qy 422 M|E|T|A|T|R|I|L|E|P|R|O|L|E|U|A|S|G|I|N|G|I|N|U|L|Y|S|G|I|L|Y|G|A|A|A|I|L|E|A|S|P|A|I 441
Db 63 A|T|G|C|A|T|A|C|A|T|T|C|T|T|G|A|T|G|A|G|A|T|G|A|A|G|G|T|A|A|A|G|C|G|C|T|A|T|T|G|A|T|T 122
Qy 442 V|A|I|A|A|A|A|A|G|P|H|E|G|I|L|Y|S|E|R|G|Y|G|L|A|L|E|U|I|E|G|I|N|G|I|L|E|U|L|E|U|A|I|L|E 461
Db 442 V|A|I|A|A|A|A|A|G|P|H|E|G|I|L|Y|S|E|R|G|Y|G|L|A|L|E|U|I|E|G|I|N|G|I|L|E|U|L|E|U|A|I|L|E 461
Qy 123 G|T|G|T|G|C|A|A|C|C|C|T|T|G|G|G|A|A|T|C|G|A|G|G|T|C|T|G|A|T|C|C|G|A|G|G|T|T|A|T|A|T|C|T|G|C|A 182
Qy 462 C|Y|S|G|I|S|E|R|I|E|G|Y|A|L|E|U|T|H|R|P|R|O|Y|L|E|U|A|I|A|L|E|U 476
Db 183 T|T|C|G|T|T|C|T|C|T|G|G|A|A|C|T|G|A|C|A|C|C|T|T|G|G|G|G|A|T|A|T|C 227
```

```
RESULT 4
US-09-313-294A-3344
; Sequence 3344, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 3344
; LENGTH: 265
; TYPE: DNA
; ORGANISM: Zea mays
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700611386H1
; NAME/KEY: unsure
; LOCATION: 28, 234
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-3344

Alignment Scores:
Pred. No.: 2.3e-11 Length: 265
Score: 168.00 Matches: 34
Percent Similarity: 76.92% Conservative: 6
Best Local Similarity: 65.38% Mismatches: 12
Query Match: 6.39% Indels: 0
Gaps: 0

US-09-869-433-2 (1-515) x US-09-313-294A-3344 (1-265)
Qy 421 G|U|U|E|R|A|T|R|I|P|R|O|L|E|U|A|S|G|I|N|G|U|L|Y|S|G|I|L|Y|S|A|A|A|I|L|E|A|S|P 440
Db 106 GAAATGCAATATATTCCTTTGATGAGATATGAGATTAAGGAAAGGAGCCATTGAT 165
Qy 441 V|A|I|A|A|A|A|A|G|P|H|E|G|I|L|Y|S|E|R|G|Y|G|L|A|L|E|U|I|E|G|I|N|G|I|L|E|U|L|E|U|A|I|L|E 460
Db 166 G|T|G|T|T|G|C|A|A|C|C|A|C|T|G|G|G|C|A|A|G|T|G|G|G|G|G|C|A|T|T|T|C|C|A|C|A|G|T|T|C|A|T|C|T|G 225
Qy 461 I|L|E|C|Y|S|E|R|I|E|G|Y|A|L|E|U|T|H|R|P|R|O|Y|L|E|U 472
Db 226 A|C|G|T|T|C|G|A|C|C|C|T|G|G|C|A|A|T|G|C|A|T|C|C|A|T|C|T 261

RESULT 5
US-09-252-991A-15706/C
; Sequence 15706, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15706
; LENGTH: 1365
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15706

Alignment Scores:
Pred. No.: 3.44e-06 Length: 1365
```

Score: 135.00 Matches: 110
 Percent Similarity: 32.25% Conservative: 58
 Best Local Similarity: 21.11% Mismatches: 163
 Query Match: 5.13% Indels: 190
 Gaps: 26

US-09-869-433-2 (1-515) x US-09-252-991A-15706 (1-1365)

Qy 32 LeuMetPhePheCysIleThrPheAsnTyrThrValLeuArgAspThrLysAspThrLeu 51
 Db 1248 CTGGTGTCTTCCTGGACACCTGAACTACATC-----GATCGCGCC 1207
 Qy 52 IleValGly-----AlaProGlySerGlyAlaGluAlaIleProPhe 65
 Db 1206 GCCCTCGGCGGTATGACCGCATCTTGCAGAAAGATGAGCTGAGCGCATGAGCTTAC 1147
 Qy 66 -----IleLysPheThrPheValValProCysAlaIleIlePheMetLeuIleTyrAla 83
 Db 1146 GCCAACATCACTCTGCTTCCAGTCCGCTACGCCATCGGCTTCATCTCCAGGCGCC 1087
 Qy 84 LysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThrProPheLeu 103
 Db 1086 TTCATTCACAAAGCGCGGCTCAAGCGCGCTTCTCTCGCGCTG----- 1042
 Qy 104 IlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeuHisProThr 123
 Db 1042 ----- 1042
 Qy 124 GluPheAlaAspArgLeuGlnAlaIleLeuProGlyLeuLeuGlyLeuValAlaIle 143
 Db 1041 -----CTGCTCTGGAGCTGGCCACCGCGCCGCCAGCGCTGGCCAGCCTCG 997
 Qy 144 LeuArgAsnThrPheIleAlaIlePheTyrVal-----LeuAlaGluLeu-----TrpGly 160
 Db 996 GCGCGCGGCTCATGCTGTGCTGCTCATCTCGGCTTCAAGCGAGGCGGCCCATACCG 937
 Qy 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaGlnGluIleThrLysIleHisGlu 180
 Db 936 GCTCGCTGAAGACACCCCGCTGTGTCCCGCGGGGAA----- 895
 Qy 181 AlaLysArgPheTyrAla-----LeuPheGlyIleGlyAlaAsnIleSerLeuLeuAla 198
 Db 894 -----CGGGCGGTGGCCACCGCATCTTCAAGCCGGCACATGTCGGCGCATGTC 841
 Qy 199 SerGlyArgAlaIle-----ValThrAlaSerLysLeuArgAlaSerVal 213
 Db 840 ACCCGCGCCTGTGCTGCTGATCTCAGCGCTCGGCGC----- 802
 Qy 214 SerGluGlyValAspProTrpGlyIleSerLeuArgLeuMetAlaMetThrIleVal 233
 Db 801 -----TGGCAGGCGCGTTCTCTTGCATGTCGGCGCTG----- 769
 Qy 234 SerGlyLeuValLeuMetAlaSerTyrTrpThrIleAsnLysAsnValLeuThrAspPro 253
 Db 768 -----GGCGTGGCTGGCGGTG-----TTCGTGGTGGCGAAC----- 736
 Qy 254 ArgPheTyrAsnProGluGluMetGlnLysGlyLysLysGly----- 267
 Db 735 -----TACTAACACCCCGAGGATCCCGGCGTGAAGACGACGACGATGATATCAG 679
 Qy 268 -----AlaLysProLysMetAsnMetLysAspSerPheLeuTyrLysAspArgSer 284
 Db 678 CAGCAGAGAGAGCGCGCCGCCACCGCGGTGCTTCTGAGATTCGCCCGCGCGCGC 619
 Qy 285 ProTyrIleLeuLeuLeuThrIleValIleAla----- 296
 Db 618 ACCTGGGCGCTTGGCGCTGCTACTCATCACCGCGCGGTCTGTGTTCTACTCTAC 559
 Qy 297 -----TyrGlyIleCysIleAsnLeuIleGluValThr 307
 Db 558 TGGCTGCGCGCTTCTCAACACGACAGTACGCGCTGGCGCATCAAGCGACGATG-- 502
 Qy 308 TrpLysSerGlnLeuLysLeuGlnTyrProAsnMetAsnAspTyrSerGluPheMetGly 327

Db 502 ----- 502
 Qy 328 AsnPheSerPheThrPheGlyValValSerValLeuIleMetLeu----- 342
 Db 501 -----GGCATCCCGCTGATCTGATCTGTTGACCCGCGCATCTTGGC 460
 Qy 343 PheValGlyLysAsnValIleArgLysPheGlyTyrPheThrGlyAlaLeuValThrPro 362
 Db 459 AGCGTGGCGCGGACATCTTTCG-----TCTGGCTGATCGGCGCGCATCGCGC 406
 Qy 363 ValMetValLeuLeuThrGlyIleValIlePhePheAlaLeu-----ValIlePhe 378
 Db 405 ATCCGCGCGCGCTGCTGATGATGCTGCTTGGCCGCGACCATCGCGCGTGTTC 346
 Qy 379 ArgAsnGlnAlaSerGlyLeu-----ValAlaMetPheGlyThrThrProLeuMetLeuAla 397
 Db 345 GCCGCCAACCGCACGCGCTGTGATCGCGCTG-----CTGGCCATCGCC 301
 Qy 398 ValValValGlyAlaIleGln-----AsnIleLeuSerLysSerThrLysTyr 413
 Db 300 CTGGCGGTGGCGCGCACGCGGTGACGCGCAACATCTGAGC----- 256
 Qy 414 AlaLeuPheAspSerThrLysGluMetAlaTyrIleProLeuAspGlnGlnLysVal 433
 Db 255 -----CTGTGATGACTACACGCCCAAGCACCATGAGACGCTG 214
 Qy 434 LysGlyLysAlaAlaIleAspValValAlaAlaArgPheGly-----LysSer 449
 Db 213 TTCGGC-----TTCGCGCGCATGTGCGCGCGCTG 184
 Qy 450 GlyGlyAlaLeuIleGlnGlnGlyLeuValIleLysGlySerIleGlyAlaMetThr 469
 Db 183 GCGGGATGTTCAAGACCG-----ATGCTCGCGCGCTACTGACCGCGACGAAC 133
 Qy 470 ProTyrLeuAlaValIleLeuLeuPheIleAlaIleThrLeuValSerAlaThrLys 489
 Db 132 AACACATCCGCGGTCTTCCACCATGATCCGCGCATGATCTTCATCGCGCTGACCTGG 73
 Qy 490 Leu 490
 Db 72 CTG 70

RESULT 6

US-09-252-991A-15648
 ; Sequence 15648, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 15648
 ; LENGTH: 1638
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-15648

Alignment Scores:

Pred. No.: 4 6e-06 Length: 1638
 Score: 135.00 Matches: 110
 Percent Similarity: 32.25% Conservative: 58
 Best Local Similarity: 21.11% Mismatches: 163
 Query Match: 5.13% Indels: 190
 Gaps: 26


```

Qy 52 IleValGly-----AlaProGlySerGlyAlaGluAlaIleProHe 65
Db 359 GCCCTGGCGTATGATGACGCCATCTCTGCCGAGAAAGATGAGTGCATGACATGACATAC 418
Qy 66 -----IleLysPheTrpLeuValValProCysAlaIleIlePheMetLeuIleTyrAla 83
Db 419 GCCAACATCAACTTCTGTTCCAGGCTTCGCGCTACGCGCATCGCTTCATCTCCAGAGGCGCG 478
Qy 84 LysLeuSerAsnIleLeuSerTyrGlnAlaLeuPheTyrAlaValGlyThrProPheLeu 103
Db 479 TTCATCGACAGAGTGGGGTCAAGCGCGCTTCTCTCGCGCTG----- 523
Qy 104 IlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValIleHisProThr 123
Db 523 ----- 523
Qy 124 GluPheAlaAspArgLeuGlnAlaIleLeuProGlyLeuLeuGlyLeuValAlaIle 143
Db 524 -----CTGCTCTGAGCCTGGCCACCGGCCACGCGCTGGCGACTCG 568
Qy 144 LeuArgAsnTrpThrPheAlaAlaPheTyrVal-----LeuAlaGluLeu-----TrpGly 160
Db 569 GCGGCGGCTTCATGCTGTCTGCTTCATCTCGCGCTGACCGAGCGCGCAACTACCGG 628
Qy 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrTyrIleHisGlu 180
Db 629 GCCTCCGAGAACACACCGCGCTGTGCTTCCCGCGGGGAA----- 670
Qy 181 AlaLysArgPheTyrAla-----LeuPheGlyIleGlyAlaAsnIleSerLeuLeuAla 198
Db 671 -----CGGCGGCTGGCCACCGCATCTTCACAGCGGCACCAATGTCGGCGCATGGTC 724
Qy 199 SerGlyArgAlaIle-----ValTrpAlaSerLysLeuArgAlaSerVal 213
Db 725 ACCCGCGCTGTTCCTGCTATCTCATCGCTCGGCGC----- 763
Qy 214 SerGluGlyValAspProTrpGlyIleSerLeuArgLeuLeuMetAlaMetThrIleVal 233
Db 764 -----TGGCAGCGCGCTTCTCTCTGATGTCGCGCGCTG----- 796
Qy 234 SerGlyLeuValLeuMetAlaSerTyrTrpIleAsnValAsnValLeuThrAspPro 253
Db 797 ---GGCGTGTCTGCTGCTGTG---TTCTGTGTGGCGCAAC----- 829
Qy 254 ArgPheTyrAsnProGluGluMetGlnLysGlyLysGly----- 267
Db 830 ---TACTACAACCCGAGAGCATCCGCGGTGAAGACAGAGCGAGCTGAAATACATCCAG 886
Qy 268 -----AlaLysProLysMetAsnMetLysAspSerPheLeuTyrLeuAspArgSer 284
Db 887 CAGCAGAGAGAGCGCGCCGCCACCGCGCTTCTCTGACAGATTCTCCGCGCGCGCGC 946
Qy 285 ProTyrIleLeuLeuThrLeuLeuValIleAla----- 296
Db 947 ACCTGGGCGCTTCGCGCTGCTACATCGATCACCGCGCGGTTCGTGTTCACTTAC 1006
Qy 297 -----TyrGlyIleCysIleAsnLeuIleGluValThr 307
Db 1007 TGAGTCGCGCGCTTCTCTCAACAGCAGATACCGCTCGGCGCATACAGCGTACAGATG--- 1063
Qy 308 TrpLysSerGlnLeuLysLeuGlnIleTyrProAsnMetAsnAspTyrSerGluPheMetCly 327
Db 1063 ----- 1063
Qy 328 AsnPheSerPheTrpThrGlyValValSerValLeuIleMetLeu----- 342
Db 1064 -----GGCATCCCGGTGATCTGATCTGTGTACACCGCGCATCTTGGC 1105
Qy 343 PheValGlyGlyAsnValIleArgLysPheGlyTrpLeuThrGlyAlaLeuValThrPro 362
Db 1106 AGCGTGGCGCGCGCATCTTTCG-----TCTGTGGTGTATCGCGCGCATCCGCGCG 1159
Qy 363 ValMetValLeuLeuThrGlyIleValPhePheAlaLeu-----ValIlePhe 378

```

```

Db 1160 ATCCGCGCGCGCTGCTGATGCTGTCTGCTTCCGCTGACCATGTCGCGGTGTCTC 1219
Qy 379 ArgAsnGlnAlaSerGlyLeu---ValAlaMetPheGlyThrThrProLeuMetLeuAla 397
Db 1220 GCCGCCAACGCGACCGCGCTGTGTGATGCGCGCTG-----CTGGCATCTGCC 1264
Qy 398 ValValValGlyAlaIleGln-----AsnIleLeuSerLysSerThrLysTyr 413
Db 1265 CTGGCGGTTCGCGCGCACACGCGTGCAGCGCGCAACTGTGAGC----- 1309
Qy 414 AlaLeuPheAspSerThrLysGluMetAlaTyrIleProLeuAspGlnGluLysVal 433
Db 1310 -----CTGGTATGAGACTACACGCCCAAGCACTGATGACGCGTGTG 1351
Qy 434 LysGlyLysAlaAlaIleAspValValAlaAlaArgPheGly-----LysSer 449
Db 1352 TTCGCG----- 1381
Qy 450 GlyGlyAlaLeuIleGlnGlnGlyLeuLeuValIleCysGlySerIleGlyAlaMetThr 469
Db 1382 GCGCGGATGTTCAATGACCCAG-----ATGTCGCGCGGCTGACTGACCGCAGCAGAC 1432
Qy 470 ProTyrLeuAlaValIleLeuLeuPheIleIleAlaIleTrpLeuValSerAlaThrLys 489
Db 1433 AACACATACGCGGCTGCTTACCATGATCCGCGCATGTACTTCAATGCTTCACTTGG 1492
Qy 490 Leu 490
Db 1493 CTG 1495

RESULT 8
US-09-557-884-1/C
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557, 884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 183021 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

```

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Alignment Scores:
Pred. No.: 0.554 Length: 1830121
Score: 133.00 Matches: 120
Percent Similarity: 35.42% Conservative: 89
Best Local Similarity: 20.34% Mismatches: 158
Query Match: 5.06% Indels: 224
DB: 4 Gaps: 30

US-09-869-433-2 (1-515) x US-09-557-884-1 (1-1830121)

```

QY 15 PheLeuTrpProIleHisThrHisGluLeu-----LysIleValLeuPro 29
DB 1779409 TTTATTTGCGCCAGACCAAACTGACATTATGAAATCAAGCTAAAGTGGTATTTTGCT 1779350
QY 30 MetPheLeuMetPhePheCysIleThrPheAsnYrThrValLeuArgAspThrIleAsp 49
DB 1779349 AATTAGTGGTTTATGTTACTTACTTTCAGTG----- 1779314
QY 50 ThrLeuIleValIleGlyAlaProGlySerGlyAlaGluAlaIleProPheIleLysPheTrp 69
DB 1779313 -----TTTTCAGATTTTTAA 1779299
QY 70 LeuValIleProCysAlaIleIlePheMetLeuIleTyRAlaLysLeuSerAsnIle--- 88
DB 1779298 TTTAATT-----TTATCAGTAAAGTCTTGGCAATATAAAA 1779263
QY 89 -----LeuSerLysGlnAlaLeuPheTyR 96
DB 1779262 CTAGGCGCAAGATGAAGAAGAACCTGATTTACTTCTTATCTTGCGCTTGCATGTTATTT 1779203
QY 97 AlaValIleGlyThrProPhe---LeuIlePhePheAlaLeuPheProThrValIleTyRPro 115
DB 1779202 GCCCGTGGATGGGGGTTGGCTGATGTTTTTGGCGTAGCAAACTTAACCATTAAT 1779143
QY 116 LeuAspArgValLeuHisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProPro 135
DB 1779142 CTTTCTGACATTACAGAGGTTCTGCAGAACCAAACTTAACCATTAAT----- 1779089
QY 136 GlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTyRValLeu 155
DB 1779088 -----CATACTCTGTTCCACTGGGAATTCACGCGGCGGACATTA--- 1779050
QY 156 AlaGluLeuTrpGlySerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIle 175
DB 1779049 -----TATGGACCATTCCTTTAGCATTAAGCTTATTTGGGTTTCGTTATAAATTA 1778999
QY 176 Thr-----LysIle 178
DB 1778998 CCTTAGCGGTGGCTCTGTTGTTTAACTTTATTAAGATCGTAATTAAGGCAAAATTC 1778939
QY 179 HisGluAlaLysArgPheTyRAla-----LeuPheGlyIle----- 190
DB 1778938 GGGCATTCGCAATTGATGATTAGCGCTTACTTGCACATTAATTTGGGATTAATTAACCATTA 1778879
QY 191 GlyAlaAsnIleSerLeuLeuAlaSerGly--ArgAlaIleValIleTyRAlaSerLysLeu 209
DB 1778878 GGTTTTGGTTCATACAGCTTGTCAGGGCTTGAACAAATGAGTTGATTAAGTCAGAAC 1778819
QY 210 ArgAlaSerValSerGluGlyValAspProTrpGlyIleSerLeuArgLeuMetAla 229
DB 1778818 AGCTTTGCTTACCAAGTTGGCGTTATTT-----GTTTGGTATGATGTTTACGCGGTG 1778768
QY 230 MetThrIleValSerGlyLeuValIleMetAlaSerTyYrTrpIleAsnLysAsnVal 249
DB 1778767 TTTTCTGCTATTTCTGCTGTT----- 1778747
QY 250 LeuThrAspProArgPheTyRAsnProGluGluMetGlnLysGlyLysIleGlyAlaLys 269
DB 1778746 -----GGAAGGCGGTGAAA 1778732

```

```

QY 270 -----ProLysMetAsnMetLysAspSer-PheLeuTyR----- 280
DB 1778731 ATATTAGTGAATTAATTAACCTTAACCTTAACCTTATTTACTTCTTCTTTTAATTCA 1778672
QY 281 ----LeuAspArgSerProTyRLeuLeuLeuThrLeuValIleAlaTyRGlYI 299
DB 1778671 GGCCCTACGATATACCTTTATTCGATTTAGCACAATATTTGGTAATT-----AT 1778621
QY 299 eCysIleAsnIleGluValIleThrPlySerGlnLeuLysLeuGlnTyRProAsnMe 319
DB 1778620 TTCAGC-AATTTAGTGCACACAGTTTAAAC-----TATGCTTAAGA 1778577
QY 319 cAsnAspTySerGluPheMetGlyAsnPheSerPhe----- 331
DB 1778576 ACAAGAAATATCTAGCTGTTTACCGGATGACGTGCTTATTTGGCGTGGTGCTTC 1778517
QY 332 -TrpThrGlyValIleSerValLeuIleMetLeuPheValGlyLysAsnValIleArgly 351
DB 1778516 TTGGGCTCGTTTGGGTTTATTTATTTAGCCGATTTCTAAAGGCGCACTATCCGTGA 1778457
QY 351 sPheGlyYrTrpLeuThrGlyAlaLeuValIleProValMetValLeuLeuThrGlyIleVa 371
DB 1778456 ATTT-----ATTTTGGCGTATTAATTTCA-----AGTTTATTTGGTATTTT 1778412
QY 371 lPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyTh 391
DB 1778411 ATGGTTT-----ACTGTTTTCGTA 1778391
QY 391 rThrProLeuMetLeu-----AlaValIleGlyAlaIleGlnAsnIleLeuSerly 409
DB 1778390 TACAGAGATATGCTAAATATGATGACATTCGCGGCGGCGGCGGCAATTAATTTCT-- 1778333
QY 409 sSerThrIleTyRAlaLeuPheAspSerThrIleGluMetAlaTyRleProLeu----- 427
DB 1778332 -TCCCGAAGAAATTTATTAATTT-----AAATTTTAATTAATTCGCTTTTACCAAC 1778283
QY 427 ----- 427
DB 1778282 AATTAACAGCTTTGTGACCTTATTAAGTATTTATTTTATCACTTCAGCGGATTC 1778223
QY 428 -----AspGlnGlu----- 430
DB 1778222 AGTATTATATGTTAAATAATACATTGCATCTCGATTAAGTTTAACTTCGCTGCGTG 1778163
QY 431 -GlnLysValIleGlyLysAlaAlaIleAspValAlaAlaArgPheGlyLysSerG 450
DB 1778162 GCAAGCCATATATGGGGAACTTAATGTCGTTGTCGATGCTTATGATCAATCTGG 1778103
QY 450 YGly---AlaLeuIleGlnGlnGlyLeuLeuValIleCysGlySerIleGlyAlaMetTh 469
DB 1778102 TGGACTTGCTAATTTGCAACAAATGACATTATTTGTC-----TT 1778061
QY 469 rProTyLeuAlaValIleLeuLeuPheIleAlaIleTrp-----LeuValSer-- 486
DB 1778060 GCGTTTGGCTTATTAATGATGTTGAATGTTTAAAGTTTAAAGGCTTAATTCGGA 1778001
QY 487 -----AlaThrLysLeuAsn 491
DB 1778000 TAAAAAATATTTTCTACTTAAAGTCAAT 1777973

```

RESULT 9

US-09-643-990A-1/c
Sequence 1, Application US/09643990A
Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White

Hamilton O. Smith

J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville,
 STATE: MD
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3 1/2 inch diskette
 COMPUTER: Dell Pentium
 OPERATING SYSTEM: MS DOS v.6.22
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/643,990A
 FILING DATE: 23-Aug-2000
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/487,429
 FILING DATE: 1995-06-07
 APPLICATION NUMBER: 08/426,787
 FILING DATE: 1995-04-21
 ATTORNEY/AGENT INFORMATION:
 NAME: Kenley K. Hoover
 REGISTRATION NUMBER: 40,302
 REFERENCE/DOCKET NUMBER: PB186P1C1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 301-610-5790
 TELEFAX: 310-309-8439
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1830121 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:
 US-09-643-990A-1
 Alignment Scores:
 Pred. No.: 0.554 Length: 1830121
 Score: 133.00 Matches: 120
 Percent Similarity: 35.42% Conservative: 89
 Best Local Similarity: 20.34% Mismatches: 158
 Query Match: 5.06% Indels: 224
 DB: 4 Gaps: 30
 US-09-869-433-2 (1-515) x US-09-643-990A-1 (1-1830121)
 QY 15 PheLeuTrpProIleHisThrHisGluLeu-----LysLysValLeuPro 29
 DB 1779409 TTTATGGCCAGAGCAAACTTCAAGCATTTGAATCAAGTAAAGTGATTTTGGCT 1779350
 QY 30 MetPheLeuMetPhePheCysIleThrPheAsnTrpThrValLeuArgAspThrLysAsp 49
 DB 1779349 AATTTAGTGGTTTATGATCTACTTTTCAGTG----- 1779314
 QY 50 ThrLeuIleValIGlyAlaProGlySerGlyAlaGluAlaIleProPheIleLysPheTrp 69
 DB 1779313 -----TTTTTGAGATTTTGA 1779299
 QY 70 LeuValValProCysAlaIleIlePheMetLeuIleTyralAlaLysLeuSerAsnIle--- 88
 DB 1779298 TTAATTT-----TTATAGTAAAGTACTTGGCAATATAAA 1779263
 QY 89 -----LeuSerIlySGlnAlaLeuPheTyR 96
 DB 1779262 CTAGGCAAGATGAAGAAGAACTGATTTAGTTCTTATCTTGCGCTTGCGATGTATTT 1779203
 QY 97 AlaValGlyThrProPhe---LeuIlePhePheAlaLeuPheProThrValIleTyRPro 115
 DB 1779202 GCCCGTGAATGGGGGTTGGCTGATGTGTTTTTGGCGTAGAGAAACCAATTAAACCATTTAT 1779143
 QY 116 LeuArgAspValLeuHisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProPro 135

DB 1779142 CTTTCTGACATTTACAGAGGTTCTGCAGAAACATTAACAAACAGAGGCTTACTT----- 1779089
 QY 136 GlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTyRValLeu 155
 DB 1779088 -----CATACTCTGTTCCAGGGAAATTCACGGCTGGCAGTA--- 1779050
 QY 156 AlaGluLeuTrpGlySerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIle 175
 DB 1779049 -----TATGGACACATTCCTTTAGCATTTACTTATTTGGGTTTCGTTAAATTA 1778999
 QY 176 Thr-----LysIle 178
 DB 1778998 CTTTAGCGTTGGCTCTGTTTATCTTATTAAAGTCGATTAAATGCAAAATC 1778939
 QY 179 HisGluAlaLysArgPheTyRAla-----LeuPheGlyIle----- 190
 DB 1778938 GGCATGCAATTGATGTATATGCGCTTACTTGCCACATTAATTTGGATTAATACCAATTA 1778879
 QY 191 GlyAlaAsnIleSerLeuLeuAlaSerGly---ArgAlaIleValITrpAlaSerLysLeu 209
 DB 1778878 GGTTTGGTTTCATCAGACTTGCTGTCAGGCTTGCAACAATAGGTGATTAAGTCAGAAC 1778819
 QY 210 ArgAlaSerValSerGluGlyValaLysProTrpGlyIleSerLeuArgLeuMetAla 229
 DB 1778818 AGCTTGCCCTTACAAAGTTGGCGTTAT-----GTTGGTGATGTTGTTAGCGGTG 1778768
 QY 230 MetThrIleValSerGlyLeuValLeuMetAlaSerTyRTrpTrpIleAsnLysAsnVal 249
 DB 1778767 TTTTCTGCTATTTCTGCTGTT----- 1778747
 QY 250 LeuThrAspProArgPheTyRAsnProGluGluMetGlnLysGlyLysGlyAlaLys 269
 DB 1778746 -----GCAAAAGGCGTGAA 1778732
 QY 270 -----ProLysMetAsnMetLysAspSer-PheLeuTyR----- 280
 DB 1778731 ATATTAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1778672
 QY 281 ---LeuAspArgSerProTyRILELeuLeuLeuLeuLeuValIleAlaTyRGLYI 299
 DB 1778671 GGCCCTTCGTTATACCTTTATATCGATTAGCAGCAATATTTGGTAAATTT-----AT 1778621
 QY 299 eCysIleAsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGlnTyRProAsnMe 319
 DB 1778620 TTCAGC-AATTTAGTGCACCTCAGCTTTTAAAC-----TATGCTTAGA 1778577
 QY 319 tAsnAspTyRserGluPheMetGlyAsnPheSerPhe----- 331
 DB 1778576 ACAAGAACATATCTAGCTGCTTTAGCGGATGACGCTGCTTATTTGGCTTGCTGCTGTC 1778517
 QY 332 -TrpThrGlyValaLysValLeuIleMetLeuPheValIGlyLysAsnValIleArgGly 351
 DB 1778516 TTGGGCTCCGTTGTGTGTTATTTATTTAGCGCGATTTCTTAAGGGGCAACTATCCGCTGA 1778457
 QY 351 sPheGlyTrpLeuThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleVal 371
 DB 1778456 ATTT-----ATTTTGGCGATTAGTTATTC-----AGTTATTTGGTATTTT 1778412
 QY 371 lPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyTh 391
 DB 1778411 ATGCTTT-----ACTGTTTGGTAA 1778391
 QY 391 rThrProLeuMetLeu-----AlaValValaGlyAlaIleGlnAsnIleLeuSerTy 409
 DB 1778390 TACAGCAGATGATGCTTAATATGATGCGATGCTGCGGCGGCTGCGCAATTAATTTCT- 1778333
 QY 409 sSerThrLysTyRAlaLeuPheAspSerThrLysGluMetAlaTyRlIleProLeu 427
 DB 1778332 -TCCCGCAAAATTTTATTAATTT-----AAATTTTAATATATCTGCTTTAACCAAC 1778283
 QY 427 ----- 427

```

Db 1778282 AATAACAGGCTTGTGAGCTTATAGTATTTATTTTATCATCTGACGGAGATTC 1778223
Qy 428 -----AspGlnGlu----- 430
Db 1778222 AGGATTTATGTTGTTAAATACATTCCTCGTATATAAGTTAGCTGCGCTGCG 1778163
Qy 431 -GlnlyVallysglyValalaIleaspValValAlaIleargPheglyLysSerGI 450
Db 1778162 GCAACCCATATGTCGGAGACTTATATGCTTTTCGATGTCGATGCTTATGCAATCTGG 1778103
Qy 450 yGIY---AlaLeuIleGlnGlnGlyLeuLeuValIleCysglySerIleGlyAlaMetTh 469
Db 1778102 TGGACTTCCTATTTGCAACAACATACATTAATTTGTCG-----TT 1778061
Qy 469 rProGlyLeuAlaValIleLeuLeuPheIleIleAlaIleTrp-----LeuValSer-- 486
Db 1778060 GCCTTTTCCTTATGATGTTGATATGTTTATGTTATGAAAGGCTTAATGCGGA 1778001
Qy 487 -----AlaThrLysLeuAsn 491
Db 1778000 TAAAAAATATTTTCTACTAAAGTCAAT 1777973

RESULT 10
US-09-252-991A-1331
; Sequence 1331, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Mate J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1331
; LENGTH: 1467
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1331

Alignment Scores:
Pred. No.: 2.01e-05 Length: 1467
Score: 129.00 Matches: 111
Percent Similarity: 32.62% Conservative: 71
Best Local Similarity: 19.89% Mismatches: 180
Query Match: 4.90% Indels: 196
Gaps: 23

US-09-869-433-2 (1-515) x US-09-252-991A-1331 (1-1467)
Qy 71 ValValProCysAlaIleIle-----PheMetLeuIleTrpAlaLysLeuSer 86
Db 7 GTCAACCCCTGCGCGCTCGCGGCGGAGGTCGTGTCGGGTCGCGGATTC 66
Qy 87 AsnIleLeuSerLys----- 91
Db 67 CGCGTCCTGCTCCGCGGCGCTCCCGGAGCTCAACGCTGTGCGAGCGGCGAGTCCGGA 126
Qy 92 -----GlnAlaLeuPheTrpAlaValGlyThrPro-----PheLeuIlePhe 105
Db 127 GCGGCGCGCGCTCCCTGTACGAGACCGAGGTCCGATGCGCGGATTTTCTCTCTTT 186
Qy 106 PheAlaLeuPheProThrValIleTrpProLeuArgSerValIleuHisProThrGluPhe 125
Db 187 CGGCTCTCTAT-----TTC 201
Qy 126 AlaAspArgLeuGlnAlaIleLeuProGlyLeuLeuGlyLeuValAlaIleLeuArg 145
Db 202 GCTACCGCTGCTGAG--CTACACGGGCTCGGCGCTGCTGACG----- 240

```

```

Qy 146 AsnTrpThrPheAlaIlePheTrpValIleuAlaGlu-----LeuTrpGlySer 161
Db 241 -----ACCTACCTGGGCGCTGCGCGCGGACCAAGTCAAGCGCCTGTGGGTGGC 294
Qy 162 ValMetLeuSerLeuMetPheTrpGlyPheAla-----AsnGlu 174
Db 295 GCGGTATAGGCGGCAACTACTTGGCTGGTGTGCGGCGGCAAGTCCGATCACGGCTG 354
Qy 175 IleThrLysIleHisGlnAlaLysArgPheTrpAlaLeuPheGlyIleGlyAlaAsnIle 194
Db 355 ATCCCGCGGCGCGCACATCCGCGCTACGTCGCTGCGCGCGGGGTG----- 402
Qy 195 SerLeuLeuLaseSerGlyArgAlaIleValTrpLaseSerLysLeuArgAlaSerValSer 214
Db 403 -----GTCACCGCGCGGCTCTGGC 423
Qy 215 GlnGlyValAspProTrpGlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSer 234
Db 424 CATGGCTGCGCGCTGG-----CTGCGGCGCTGATTCGCTTGGCGATGATG 474
Qy 235 GlyLeuValLeuMetAlaSerTrp-----TrpIleAsnLysAsnValIleu 250
Db 475 GGCCTGGCGCTGATGTCCAGTACATGATGATGAGAGCTGCGTCAAC----- 522
Qy 251 ThrAspProArgPheTrpAsnProGlnGluMetGlnLysGlyValAlaLysPro 270
Db 523 -----GACACAGCGCGATCCACGCCGCGGCGCGCTTTC 558
Qy 271 LysMetAsnMetLysAspSerPheLeuTrpLysAspArgSerProTrpIleLeu----- 288
Db 559 GCGCGCTACATGCGCGCTTCTACTGCGCTGGTGTGCGGCGGAGATCTGTGGGCG 618
Qy 289 -----LeuLeuThrLeuLeuValIleAlaTrpGlyIleCysIle 301
Db 619 CATCCGACATGCGGCGGAGATGCTGATGCTGTGCGCTTCTGCTGCGCTGTGCTG 678
Qy 302 AsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGlnTrpProAsnMetAsnAsp 321
Db 679 GTGCGCTGCGCTGACCCACACAGATCCACCCGACGCTGCGCGCGCGCGCTGAA 738
Qy 322 TyrSerGluPheMetGlyAsnPhe-----SerPheTrpThrGlyValIleSerValLeu 339
Db 739 CCACGCTTCTTACATCCGCGGCGGAGTCCGTACACACCGATATGCTCGGGGCTG 798
Qy 340 IleMet----- 341
Db 799 GTGGTCCGCTCTTCTATGCGCTGCGCGCTGTATGCCAACAGTTGGGTTGCCAAC 858
Qy 342 -----LeuPheValGlyGlyAsnValIleArgLys 351
Db 859 GAGCAGGTGCGCTGTACATGAGGCGGCTGCATCTTCCCGGCGCTGCTGTGACTGGCCG 918
Qy 352 PheGlyTrpLeuThrGlyAla-----LeuValThrProValMetValLeu 366
Db 919 CTGGCGTGTGTCGACCGCTCGGACCGGCGCTGTATTCGCGCGCTGCGCATCTG 978
Qy 367 LeuThrGlyIleValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuVal 386
Db 979 CTG-----TGCCTGTGCGCTTGGCGCTG----- 1002
Qy 387 AlaMetPheGlyThrTrpProLeuMetLeuAlaValAlaGlyAlaIleGlnAsnIle 406
Db 1003 GCATGTGTCACACAGATCCGCTGCGCTTCTGCTGCGCTGCGGATTCGCGGAGCATG 1062
Qy 407 LeuSerLysSerThrLysTrpAlaLeuPheAspSerThrLysGluMetAlaTrpIlePro 426
Db 1063 TTG-----CAGTTCACCTCTATCCGCTGACATGGCGCTTTCACAGACAT 1110
Qy 427 LeuAspGlnGlnLysValLysGlyLysAlaAlaIleAspVal-----ValAla 443
Db 1111 GTGAGACCGACGAGCGGTTTGTGCTGACCGCATGCTGCTGTGATCCTTGCGGTGCGC 1170

```

QY 444 AlaArgPheGlyLys---SerGlyGlyAlaLeuIleGlnGlnGlyLeuLeuValIleCys 462
Db 1171 GGCTGATCGGGCCGCTGGCCGCGCGCTGATGATCGG----- 1209
QY 463 GlySerIleGlyAlaMetThrProTyrlleuAlaValIleLeuLeuPheIleAlaIle 482
Db 1210 ---CTGTTGGCGCTAACATGCTATATGCTTGCATGAGCCCTGGCGCTGATCTGTGC 1266
QY 483 ThrLeuValSerAlaThrIleuAnuIleuPhe----- 494
Db 1267 TGGCGGCTCATCCGAGAAAGTCAAGCGGCTGCACCGGCTGCACGCGCGCTGCAC 1326
QY 494 ----- 494
Db 1327 CACGTGCGACCCCGGACAAATGACACGACTCCCGCTGTGCGCGCTGCATCCGCG 1386
QY 495 LeuAlaGlnSerAlaLeuLysGlnGlnGlnValAlaGlnGlnAspSerAlaPro 512
Db 1387 GTAGACGAACAGCGGTCAGAGAGAGATGTGATGTGCGGACCGACGCGCGC 1440

RESULT 11
US-09-335-409-1
Sequence 1, Application US/09335409
Patent No. 6121029
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Geerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/335,409
CURRENT FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-335-409-1

Alignment Scores:
Pred. No.: 0.0543 Length: 68750
Score: 122.50 Matches: 105
Percent Similarity: 34.68% Conservative: 67
Best Local Similarity: 21.17% Mismatches: 165
Query Match: 4.66% Indels: 159
DB: 3 Gaps: 21

US-09-869-433-2 (1-515) x US-09-335-409-1 (1-68750)
QY 76 IleIlePheMetLeuIleTyrlaLysLeuSerAmIleLeuSerIleGln----- 92
Db 3460 CTCGGGCTCATCTCGTACCGCGCGCTCGCGAGCTCGCGCGGCGGTGGCCAG 3519
QY 93 -----AlaLeuPheTyrlaValGlyThrProPheLeuIlePhePheAla 107
Db 3520 CCCGAGTGTCTCGGAGAGCTCTTCGCGCGGCTC-----GTG 3555
QY 108 LeuPheProThrValIleTyrlProLeuArgAspValLeuHisProThrGlnPheAlaAsp 127
Db 3556 CTGGGCGCTCGATCGGCGCGCTCGCGCGCGGCTTCATCGAGCCCTTTCAGAGAG 3615
QY 128 ArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrp 147
Db 3616 CCGGCGGTGCGGCTGTCTCTCGGCGATC----- 3645
QY 148 ThrPheAlaAlaPheTyrlValLeuAlaGlnLeuTrpGlySerValMetLeuSerLeuMet 167
Db 3646 -----TCCTGATAGCGCGCTCTCTCTGCTGCTG 3675

QY 168 PheTrpGlyPheAlaAsnGlnIleThrLysIleHisGlnAlaLysArgPheTyrlaLeu 187
Db 3676 ATGGCGGCGATCGAGCTGAGCTGAGCTCTCGCGAAGAGAGCGCGCGCGGCTC 3735
QY 188 PheGlyIleGlyAlaAsnIleSerIleuAlaSerGlyArgAlaIleValTrpAlaSer 207
Db 3736 TCGGCGCTCGGCGGATCGCGCCCGCTCGCGCGGCGCGCGC----- 3780
QY 208 LysLeuAlaAlaSerValSerGlnGlyValAspProTrpGlyIleSerLeuArgLeu 227
Db 3781 ---TTCGCGGCTCGTGTCTGATCGCGCCCTTCGAGCGGCTCTTCTCGAGATCGG 3837
QY 228 MetAlaMetThrIleValSerGlyLeuVal-----LeuMet 239
Db 3838 CTCTCGGATGCGCGCTGATCGGATCGCGAAGCTGATCGAGCGGATCGATCGGCG 3897
QY 240 AlaSerTyrlTrpTrpIleAsn-----LysAsnValLeuThrAspProArgPheTyrlAsn 257
Db 3898 CGCAGCTATGCGGAGTGACGCTCGCGCGGCGGCGTGCAGAGTGTGCTGCGGCTG 3957
QY 258 ProGlnGlnMetGlnGlyLysGlyAlaLysProLysMetAsnMetLysAspSer 277
Db 3958 CTCTGCGGATGACGCTGTCAGCTACGCGCGCTGCGCGCGGCTGCGCGCGCGG 4017
QY 278 PheLeuTyrlLeuAspArgSerProTyrlleuLeuLeuThrLeuValIleAlaTrp 297
Db 4018 GCGCTCTCG-----GCGAGCGGATCTCTGCTGTTCAATGAGCTCTGTC----- 4059
QY 298 GlyIleCysIleAsnLeuIleGlnValThrTrp-----Lys 309
Db 4060 GGGCGCGCGCTACCCACCTCGCATGCGCGTGGGCGCGAGCGGCGGCTCTCAAG 4119
QY 310 SerGlnLeuLysLeuGlnTyrlProAsnMetAsnAspTyrlSerGlnPheMetGlyAsnPhe 329
Db 4120 GCACAGGTGCTGCTC----- 4134
QY 330 SerPheTrpThrGlyValIleValSerValIleMetLeuPheValGlyLysAlaIle 349
Db 4135 -----GTCTCTGCTCTACAGTCTCTGCGCGCGGCTGAGC 4170
QY 350 ArgLysPheGlyTrp-----LeuThrGlyAlaLeuValThrProValMetVal----- 365
Db 4171 CAGCGGCTCGGCTGACCGCGCTGCTGCGCGCGCTGCGGCTGCTGCTGCTCAACAGC 4230
QY 366 -----LeuLeuThrGlyIle----- 370
Db 4231 GCTCTCGACCAACCGCGCTCTCTCTCGACGCGCGAGACGCTGTCGCGGCTCTTC 4290
QY 371 -----ValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeu-----Val 386
Db 4291 GCGCTGTGTCTCTGCTCTCGCGGCGATGCGCGCTGACGCTGCGAGCTGCGACGCGC 4350
QY 387 AlaMetPheGlyThrThrProLeuMetLeuAla-----ValValVal 400
Db 4351 GCGGCGTGGGAGAGGTGCGCTTCTGTCGACCGCGAGCGGCGGAGAGTGTCTCCC 4410
QY 401 GlyAlaIleGlnAsnIleLeuSerLysSerThrTyrlaLeuPheAspSerThrLys 420
Db 4411 GCGGCGCTCGCGCGCGGCTCGC-----GCGCTCAGGCGGAGCGAGCGG 4455
QY 421 GluMetAlaTyrlleProLeuAspGlnGlnLysValLysGlyLysAlaAlaIleAsp 440
Db 4456 GCGCTCGTGGGCGGTGGCGCTTAACATG-----AAGGCGGACGAGAC 4497
QY 441 ValValAlaAlaArgPheGlyLysSerGlyAlaLeuIleGlnGlnGlyLeuVal 460
Db 4498 CTCATCGTGGCGAGTGTGCG-----GTGAGCTGCGGCTCTCTC----- 4536
QY 461 IleCysGlySerIleGlyAlaMetThrProTyrlleuAlaValIleLeuLeuPheIle 480
Db 4537 -----TCAACAGAGCTTATACATGACCGCGTGTGCGGCTGCTGACGCTGAC 4587
QY 481 Ala-----IleTrpLeuValSerAlaThrLysLeuAnuLysLeuPheLeu 495

```

Db      4588 GCCTACCGCGCTCTCATCTGGCTCGAG----- 4617
Qy      496 AAGAGSerAlaLeuylsgluginlValAlaGlnGluAspSerAla 511
Db      4618 -----AAAAGGCGCCTCCGACCGCAGAGAGAGTGGCT 4650

RESULT 12
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Robb
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1

Alignment Scores:
Pred. No.:      0.0543      Length:      68750
Score:          122.50      Matches:      105
Percent Similarity: 34.68%      Conservative: 67
Best Local Similarity: 21.17%      Mismatches: 165
Query Match:      4.66%      Indels:      159
                        Gaps:      21

US-09-869-433-2 (1-515) x US-09-568-102-1 (1-68750)
Qy      76 IleIlePheMetLeuIleTyAlaLysLeuSerAsnIleLeuSerLysGln----- 92
Db      3460 CTCGGCCTATCTCTCTGACCGCGCGCTCCGGGAGCTCGCGGGCGCGCCGCG 3519
Qy      93 -----AlaLeuPheTyAlaValGlyThrProPheLeuIlePhePheAla 107
Db      3520 CCCGAGGTGCTCGGGAGCTCTTCGCGCGCGCTC-----GTG 3555
Qy      108 LeuPheProThrValIleTyProLeuArgAspValLeuHisProThrGluPheAlaAsp 127
Db      3556 CTGGGCGCTCTCGCTCGCGCGCTCGCGCGCGCGGGGTTCAGACCTCTTCCAGAG 3615
Qy      128 ArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrp 147
Db      3616 CCGGCGCTCGGGGTCTCTCTCGGCATC----- 3645
Qy      148 ThrPheAlaAlaPheTyValLeuAlaGluLeuTrpGlySerValMetLeuSerLeuMet 167
Db      3646 -----TCCTGATAGAGCGCGCTCTCTCTGCTGCTG 3675
Qy      168 PheTrpGlyPheAlaAsnGluIleThrLysIleHisGluAlaLysArgPheTyAlaLeu 187
Db      3676 ATGGCGCGATCGAGGTTCAGCGTGGGCGATCTCGCGCAGAGAGCGCGCCCGGCGCTC 3735
Qy      188 PheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGlyArgAlaIleValTrpAlaSer 207
Db      3736 TCGGCGCTCGCGCGCATCGCGCCCGCTCGCGCGCGCGCGCC----- 3780
Qy      208 LysLeuArgAlaSerValSerGluGlyValAspProTrpGlyLysSerLeuArgLeu 227
Db      3781 ---TTCTCGGCGCTCGTCTCATGATCGGCGCCCTTCGAGGCGGCGCTTCTTCGAGATCG 3837

```

```

Qy      228 MetAlaMetThrIleValSerGlyLeuVal-----LeuMet 239
Db      3838 CTCCTGATGAGCGCGCTCATCGGTATCCGAAGTCTGATCGACCGGAGTCAATCGCC 3897
Qy      240 AlaSerTyTrpTrpIleAsn-----LysAsnValLeuThrAspProArgPheTyAsn 257
Db      3898 CGCAGCTATGCGCAGATGACGCTCGCGGGGGGTGTACAGAGTCTGCTGCTGCGGTG 3957
Qy      258 ProGluGluMetGlnLysGlyLysAlaLysProLysMetAsnMetLysAspSer 277
Db      3958 CTCGTGCGATGACCTGCTCGAGCTACGCGCGCTGCGCGCGCGTGGCGCCGAGC 4017
Qy      278 PheLeuTyTrpLeuAspArgSerProTyTrpLeuLeuLeuThrLeuValIleAlaTy 297
Db      4018 GCGCTCTCG-----GCCAGCGATCTCTGCTGTTCAATGCTCTGCTC----- 4059
Qy      298 GlyIleCysIleAsnLeuIleGluValThrTrp-----Lys 309
Db      4060 GGGCGCGCGCTCACCCACCTCGCATGCGGTGGTGGCGGACGACGCGCGCTTCCAA 4119
Qy      310 SerGlnLeuLysLeuGlnTrpProAsnMetAsnAspTyTrpSerGluPheMetGlyAsn 329
Db      4120 GGACAGGTGTGCTC----- 4134
Qy      330 SerPheTrpThrGlyValValSerValLeuIleMetLeuPheValGlyLysAsnValIle 349
Db      4135 -----GTCCTCTCTCTCACGTTCTCTGCGCGCGCGCTGACG 4170
Qy      350 ArgLysPheGlyTrp-----LeuThrGlyAlaLeuValThrProValMetVal----- 365
Db      4171 CAGGCGCTCGCGCTGACCCGCTGCTGCGCGCGGTTCGCGCTCGGCTCTGCTCAAC 4230
Qy      366 -----LeuLeuThrGlyIle----- 370
Db      4231 GCTCTCGCACCAACCGCTCTCTCTCACGCGCGTGCAGACGCTGCGCGCGCTTTC 4290
Qy      371 -----ValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeu-----Val 386
Db      4291 GCGCTGTGTTCTCTCTCTGCTGCGCGCGCTGCGCGCTGACGCTGCTGCTGCTGCTG 4350
Qy      387 AlaMetPheGlyThrThrProLeuMetLeuAla-----ValValVal 440
Db      4351 GCGGCGTGGGAGCGGTGCGCTGCTGCTGCGACCGGACGCGCGCGAGGTGCTCC 4410
Qy      401 GlyAlaIleGlnAsnIleLeuSerLysSerThrTyAlaLeuPheAspSerThrLys 420
Db      4411 GCGGCGCTCGCGCGCGCTGCGC-----GGGCTCAGGGGACGAGGCG 4455
Qy      421 GluMetAlaTyTrpIleProLeuAspGlnGluGlnLysGlyLysAlaAlaIleAsp 440
Db      4456 GCGCTCGCGCGGTGGGCTTAACATG-----MAAGGCGGACGAGC 4497
Qy      441 ValValAlaAlaArgPheGlyLysSerGlyLysAlaLeuIleGlnGlnGlyLeuVal 460
Db      4498 CTCATCGTGGGATCGTGGC-----GTGAGCTCGGGCTCTCTC--- 4536
Qy      461 IleCysGlySerIleGlyAlaMetThrProTyTrpLeuAlaValIleLeuPheIle 480
Db      4537 -----TCAACAGAGCTTATACATGACCCGCTGCGCTGCTGACGCTGAC 4587
Qy      481 Ala-----IleTrpLeuValSerAlaThrLysLeuAsnLysLeuPheLeu 495
Db      4588 GCCTCACCGCGCGCTCTCATCTGGCTGAG----- 4617
Qy      496 AAGAGSerAlaLeuylsgluginlValAlaGlnGluAspSerAla 511
Db      4618 -----AAAAGGCGCCTCCGACCGCAGAGAGAGTGGCT 4650

RESULT 13
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457

```

```

GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligou, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/567,969
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-567-969-1

Alignment Scores:
Pred. No.: 0.0543 Length: 68750
Score: 122.50 Matches: 105
Percent Similarity: 34.68% Conservative: 67
Best Local Similarity: 21.17% Mismatches: 165
Query Match: 4.66% Indels: 159
Gaps: 21
DB: 4

US-09-869-433-2 (1-515) x US-09-567-969-1 (1-68750)
Qy 76 I l e l l e p h e t l e u l e u t y r a l a l y l e u s e r a n l e u s e r l y s g l n ----- 92
Db 3460 C T G C G C T C A T C C T G A C C G C G C G C C T C G G C A G C T G G C G C G C G C C A G 3519
Qy 93 ----- A l a l e u p h e t y r a l a l a l g l y t h r p h e u l e u l e p h e a l a 107
Db 3520 C C G A G G T G C T C G G G A G C T T T C G G G C G C T C ----- G T G 3555
Qy 108 L e u p h e p r o t h r v a l l e t y r p r o l e u a r g a p v a l l e u h i s p r o t h r g l u p h e a l a s p 127
Db 3556 C T G C G C C C T C C G C G T C G C G C G C G C G C G C G C G C G C G C C T T C C A G C C T T C C A G A G 3615
Qy 128 A r g l e u d i n l a l e u l e u p r o g l y l e u l e u d i l y l e u v a l a l l e l e u a r g a n t r p 147
Db 3616 C C G C G G T C G G G T C G T G C T C C G G C A T C ----- 3645
Qy 148 T h r p h e a l a l a p h e t y r v a l l e u a l a g l u l e u t r p g l y s e r v a l m e t l e u s e r l e u m e t 167
Db 3646 ----- T C C T G A N T A G C G C G C T C T C T C G C T C G 3675
Qy 168 P h e t r p l y p h e a l a a n g l u l e t h r l y l e h i s g l u a l a l y a r g p h e t y r a l a l e u 187
Db 3676 A T G C C G C G T C G A C G T C G A C G T G G C A T C C T G C G C A A G A G A G C G C C C C G G G C G C T C 3735
Qy 188 P h e g l y l l e g l a l a a n l l e s e r l e u l e u a l a s e r l y a r g a l l e v a l t r p a l a s e r 207
Db 3736 T C G C G C T C G A C G A T C G C G C C C C C G T C G G C G A G C G C G C C ----- 3780
Qy 208 L y s l e u a r g a l a s e r v a l s e r g l u g l y v a l a s p p r o t r p g l y l e s e r l e u a r g l e u 227
Db 3781 ----- T T C C G C G C G C T C G T C G A T C G G C C C T T C C A G C G C C T T C T C G G A T G T G 3837
Qy 228 M e t a l a m e t h r i l e v a l s e r g l y l e u v a l ----- L e u m e t 239
Db 3838 C T C T C G G T G A C G C G C T C A G C G T A T C G C A A G G T G C T G A T C A G C G C A G T C A T C A T C G C 3897
Qy 240 A l a s e r t y r t r p l e a n ----- L y s a n v a l l e u t h r a p p r o a r g p h e t y r a s n 257
Db 3898 C G C A G C T A T G C G A G G T G A C C T C G C G C G G G T G T G T C A C A G G T G C T G C T G G A T G 3957
Qy 258 P r o g l u m e t g l n y s g l y l y l a l a l a s p r o l y s m e t a m e t l y a a s p s e r 277

```

```

Db 3958 C T G C G C G A T G A C G T C T G A C T A C G G C G T C G C C C G C G T G G C G C C G A G C 4017
Qy 278 P h e l e u t y r l e u a s p a r g s e r p r o t r p l l e u l e u t h r l e u e u v a l l e a l a t y r 297
Db 4018 G C G C T C T C G ----- G G A C G G A T T C T T G C T C A T G A T G T C T C T C 4059
Qy 298 G l y l l e c y s i l e a n l e u l l e g l u a l t h r t r p ----- L y s 309
Db 4060 G G C G C G C G C T C A C C A C C T C G G A T C G C T G G T G G C C A C C G A C C G C G T C C A G 4119
Qy 310 S e r g l u l e u l s l e u g l n t y r p r o a m m e t a s n a s p t y s e r g l u p h e t g l y a s n p h e 329
Db 4120 G A C A G G T G C G C T C ----- 4134
Qy 330 S e r p h e t r p h r d l y a l a l s e r v a l l e u l l e m e t l e u p h e a l a g l y l a s n v a l l e 349
Db 4135 ----- G T C T C T G T C T C A C G T T C T C T G A C G C G C G C G C T G A C G 4170
Qy 350 A r g l y p h e g l y t r p ----- L e u t h n g l y a l a l e u v a l t h r p r o v a l m e t v a l ----- 365
Db 4171 C A C G G C T C G C C T G A C C C G C T G C T C G C G C G C G C G C G C G C G C T G C T C A A C A G C 4230
Qy 366 ----- L e u e u t h n g l y l e ----- 370
Db 4231 G C T C T C G C A C A C C G C C T C T C T C A C A G C G T G C A G A C G T C G G C G C C T C T T C 4290
Qy 371 ----- V a l p h e p h e a l a l e u v a l l e p h e a r a s n g l a l a s e r g l y l e u ----- V a l 386
Db 4291 G C G C T G T G T C T C T C T C G C G G C A T C G G T G C A C G T C G C A C T G C A C G C A C G C C G 4350
Qy 387 A l a m e t p h e g l y t h r t h r p r o l e u m e t l e u a ----- V a l v a l a l 400
Db 4351 G C G C G T G G G A G C G T C G C T T C T C T C A C G C G C G C A C C G C G C G A A G T C T C C C 4410
Qy 401 G l y a l a l l e g l a n l l e u s e r l y s s e r t h r l y s t y r a l a l e u p h e a s e r t h r l y s 420
Db 4411 G C G C G C T C G C G C G C G C G C T C G C ----- G G G C T C A G G C A G C A G C G C G 4455
Qy 421 G l u m e t a l a r t l e p r o l e u a s p g l n g l u l y s v a l l y s g l y l a l a l l e a s p 440
Db 4456 G C C C T C T G C G G T G G C C T A C A T G ----- A A G G C G C G C A C G A C 4497
Qy 441 V a l v a l a l a l a r g p h e g l y l y s s e r g l y l a l e u l l e g l n g l n g l y l e u v a l 460
Db 4498 C T A T C T C G C G A T C G T C G C ----- G T C A G C T C G G C T C T C 4536
Qy 461 I l e c y s l y s e r l l e g l a l a m e t h r p r o t r p l e u a l a v a l l e l e u e u p h e i l e 480
Db 4537 ----- T C C A A C G A G C T T A D A G A T A G C C G T C G C G C G T C G C A C G T G A C C 4587
Qy 481 A l a ----- I l e t r p l e u a l s e r a l t h r l y l e u d n l y l e u p h e l e u 495
Db 4588 G C C T C A C C C G C G T C T C A T C G T C G C A G ----- 4617
Qy 496 A l a g l n s e r a l a l e u l y s g l u g l n g l u a l a l a g l n g l u s p e r a l a 511
Db 4618 ----- A A A G G C G C C T C C G A C G A G A G T C G C T 4650

RESULT 14
US-09-568-480-1
Sequence 1, Application US/09568480
Patent No. 6355458
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligou, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568,480
CURRENT FILING DATE: 2000-05-10

```

| | | | |
|----|------|--|------|
| Db | 4120 | GGACAGGAGTGTCTC----- | 4134 |
| Qy | 330 | SerPheThrPthrGlyValValSerValLeuIleMetLeuPheValGlyValAsnValIle | 349 |
| Db | 4135 | -----GTCCTCTCTTCAAGTTCCTGAGCGCGCGGCTGACG | 4170 |
| Qy | 350 | ArgLysPheGlyTrp-----LeuThGlyAlaLeuValThrProValMetVal----- | 365 |
| Db | 4171 | CAGCGCTCGGCTTCACCCGCTCTCGGCGCGCTTCGCGCTGCTGCTCTACACAGC | 4230 |
| Qy | 366 | -----LeuLeuThrGlyIle----- | 370 |
| Db | 4231 | GCTCTCTGCGACCAACCGCCCTCTCTCGAGCGGCGTGACAGCTGTCGCGGCTCTTC | 4290 |
| Qy | 371 | -----ValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeu-----Val | 386 |
| Db | 4291 | CGCGCTGTGTCTTCTGCTCTCGCGGCGATGCGCTCGACGTGTGACGTGCGACGCGC | 4350 |
| Qy | 387 | AlaMetPheGlyThrThrProLeuMetLeuAla-----ValValVal | 400 |
| Db | 4351 | CGCGCGTGGGGGAGCGTCCGCTTCTCTGCGACCGCGACGCGCGGCGAGAGTGTCC | 4410 |
| Qy | 401 | GlyAlaIleGlnAlaenIleLeuSerLysSerThrLysTrpAlaLeuPheAspSerThrLys | 420 |
| Db | 4411 | CGCGCGCTCGCGCGCGCGGCTCGC-----GGCTCAAGGGGAGAGAGGCG | 4455 |
| Qy | 421 | GluMetAlaTrpIleProLeuAspGlnGlnLysValLysGlyValAlaIleAsp | 440 |
| Db | 4456 | CGCGCTCGTGGCGGCGGCTGACATG-----AAGGCGCGACACGAC | 4497 |
| Qy | 441 | ValValAlaAlaArgPheGlyLysSerGlyGlyAlaLeuIleGlnGlnLysLeuVal | 460 |
| Db | 4498 | CTCATCTCTCGGATGTCTGGC-----GTCCAGCTCGGGCTCTTC-- | 4536 |
| Qy | 461 | IleCysGlySerIleGlyAlaMetThrProTrpLeuAlaValIleLeuPheIleIle | 480 |
| Db | 4537 | -----TCCACAGAGCTTATACATGATGACCCCTCTCGGCTGCTGTCACGCTGAC | 4587 |
| Qy | 481 | Ala-----IleTrpLeuValSerAlaThrLysLeuAsnLysLeuPheLeu | 495 |
| Db | 4588 | GCCTTCACCGCGCTCTCATCTGTGGCTCGAG----- | 4617 |
| Qy | 496 | AlaGlnSerAlaLeuLysGlnGlnGlnValAlaGlnGlnLysPheAla | 511 |
| Db | 4618 | -----AAAAAGCGCGCTCCGACGACGAGAGTCTGGCT | 4650 |

RESULT 15
 US-09-568-486-1
 ; Sequence 1, Application US/09568486
 ; Patent No. 6353459
 ; GENERAL INFORMATION:
 ; APPLICANT: Schupp, Thomas
 ; APPLICANT: Ligon, James
 ; APPLICANT: Molnar, Istvan
 ; APPLICANT: Zitzke, Ross
 ; APPLICANT: Cyr, Devon
 ; APPLICANT: Goerlach, Joern
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
 ; FILE REFERENCE: 4-30582A
 ; CURRENT FILING DATE: 2000-05-10
 ; PRIOR APPLICATION NUMBER: 09/335,409
 ; PRIOR FILING DATE: 1999-06-17
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 68750
 ; TYPE: DNA
 ; ORGANISM: Sorangium cellulosum
 US-09-568-486-1

Alignment Scores:

```

Pred. No.: 0.0543 Length: 68750
Score: 122.50 Matches: 105
Percent Similarity: 34.68% Conservative: 67
Best Local Similarity: 21.17% Mismatches: 165
Query Match: 4.66% Indels: 159
DB: 4 Gaps: 21

US-09-869-433-2 (1-515) x US-09-568-486-1 (1-68750)

Qy 76 ILeIlePheMetLeuLeuTyrAlaValLeuSerAsnIleLeuSerLysGln----- 92
Db 3460 CTCGCGCTCATCTCGTACCGCGCGCTCCGCGGAGCTCGCGCGCGCGCGCTCGCGAG 3519
Qy 93 -----AlaLeuPheTyrAlaValGlyThrProPheLeuIlePhePheAla 107
Db 3520 CCCGAGGTCTCGGAGAGCTCTCGCGCGCTC-----GTG 3555
Qy 108 LeuPheProThrValIleTyrProLeuArgAspValLeuHisProThrGluPheAlaAsp 127
Db 3556 CTGGGCGCCCTCGCTCGTCCGCGCGCGCTCGCGCGCTCCATCGAGCCCTCTCCAGAG 3615
Qy 128 ArgLeuGlnAlaIleLeuProGlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrp 147
Db 3616 CCGGCGGTGCGGCGTCTCGCGCGCATC----- 3645
Qy 148 ThrPheAlaAlaPheTyrValLeuAlaGluLeuTyrPglYserValMetLeuSerLeuMet 167
Db 3646 -----TCTCGATAGCGCGCTCTCTCTGCTGCTG 3675
Qy 168 PheTrpGlyPheAlaAenGluIleThrLysIleHisGluAlaLysArgPheTyrAlaLeu 187
Db 3676 ATGGCGGCGCATCGATCGAGTGGCGCATCTCGCGAAGAGCGCGCGCGCGCGCTC 3735
Qy 188 PheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGlyArgAlaIleValTrpAlaSer 207
Db 3736 TCGGCGCTCGCGGATCGCGCCCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3780
Qy 208 LysLeuArgAlaSerValSerGluGlyValAspProTrpGlyIleSerLeuArgLeuLeu 227
Db 3781 ---TTCGCGGCGCTCGTCTCGATCGCGCCCTTCGAGCGGCGCTCTCTCGGAGATCGTG 3837
Qy 228 MetAlaMetThrIleValSerGlyLeuVal-----LeuMet 239
Db 3838 CTCCTCGGTGAGCGGCGGTCTAGCGTATCGCGAAGTCTGATCGAGCGGAGTCCATCGCC 3897
Qy 240 AlaSerTyrTrpTrpIleAsn-----LysAsnValLeuThrAspProArgPheTyrAsn 257
Db 3898 CGCAGCTATGCGCAGGTGACCTCGCGGCGGCGGTGTCAGCAGAGTCTGCTGCGGTG 3957
Qy 258 ProGluGluMetGlnLysGlyLysGlyValAlaLysProLysMetAsnMetLysAspSer 277
Db 3958 CTCGTGCGATGACGTGCTGAGCTACGCGCGCTGCGCGCGCTGCGCGCGCGCGAGC 4017
Qy 278 PheLeuTyrLeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyr 297
Db 4018 GCGCTCTCG-----GCGAGCGGATCTTCTGCTTCATGCTGCTGTC----- 4059
Qy 298 GlyIleCysIleAsnLeuIleGluValThrTrp-----Lys 309
Db 4060 GGGGCGGCGGTCCACCGACCTGCGATGCGGTGGGCGCGAGCGGAGCGCGCGCTCCAG 4119
Qy 310 SerGlnLeuLysLeuGlnTyrProAsnMetAsnAspTyrSerGluPheMetGlyAsnPhe 329
Db 4120 GGACAGGTGCTGCTC----- 4134
Qy 330 SerPheTrpThrGlyValValSerValLeuIleMetLeuPheValGlyLysAsnValIle 349
Db 4135 -----GTCTCTGCTCTCACGTTCTCTGCGCGCGCGCGCTGACG 4170
Qy 350 ArgLysPheGlyTyr-----LeuThrGlyAlaLeuValThrProValMetVal----- 365
Db 4171 CAGCGGCTCGGCTGACCGCGCTGCTGCGCGCTGCGCGCTGCGCGCTGCTGCTGCTGCTG 4230

```

```

Qy 366 -----LeuLeuThrGlyIle----- 370
Db 4231 GCTCTCGACCAACCGCCCTCTCTCTGACGGGCTGACGAGACGTCTGCGGCGCTCTTC 4290
Qy 371 -----ValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeu-----Val 386
Db 4291 GCGGCTGTGTTCTGCTGCTCGCGGCGATGCGCTGACGCTGTCGACGAGTCCGACGCGG 4350
Qy 387 AlaMetPheGlyThrThrProLeuMetLeuAla-----ValValVal 400
Db 4351 GCGGCGGTGGGAGGTGCGGCTGCTGCTGCGACCGACGCGGCGGAGGAGTGTCCCGC 4410
Qy 401 GlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLys 420
Db 4411 GCGGCGCTCGCGCGCGCGCTCGCG-----GGCTCAGGGGCGACGAGCGG 4455
Qy 421 GluMetAlaTyrIleProLeuAspGlnGluGlnLysGlyLysAlaAlaIleAsp 440
Db 4456 GCGCTCGTGGGCGGTGGGCTGGAACATG-----AAGGGCGGACGAGGAC 4497
Qy 441 ValValAlaAlaArgPheGlyLysSerGlyAlaLeuIleGlnGlnGlyLeuLeuVal 460
Db 4498 CTCATCTCGCGATCGTGGC-----GTGAGCTCGGCGCTCTCTC--- 4536
Qy 461 IleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValIleLeuLeuPheIleIle 480
Db 4537 -----TCCACGAGGCTTATGATGTAGCCGCTGCTGCGCTGTCAGGTGAC 4587
Qy 481 Ala-----IleTrpLeuValSerAlaThrLysLeuAsnLysLeuPheLeu 495
Db 4588 GCGTCAACCGCGCTCTCATCTGCTGCTGAG----- 4617
Qy 496 AlaGlnSerAlaLeuLysGluGlnGluValAlaGlnGluAspSerAla 511
Db 4618 -----AAAAGGCGCGCTCGACGAGGAGGAGTCTGCGCT 4650

```

Search completed: November 25, 2003, 13:30:51
 Job time : 2051 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 25, 2003, 10:17:18 ; Search time 5701 Seconds
(without alignments)
3695.572 Million cell updates/sec

Title: US-09-869-433-2
Perfect score: 2630
Sequence: 1 MTKTBKRFGLRSLFPIH.....AQSLKEQVNAQEDSAPASS 515

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2886711 seqs, 2045481386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV-xlh
-O=/cgn2.1/USPTO/US09869433/runat_24112003_171459_18120/app.query.fasta.1.711
-DB=genemb1 -QFMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCTL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09869433@CGN 1 1 3635 @runat_24112003_171459_18120 -NCPU=6 -ICPU=3
-NO MAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmb1.*
1: gb_da.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_srs.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mus.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_srs.*
28: em_un.*

29: em_vl.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rod.*
36: em_hcg_mam.*
37: em_hcg_vtc.*
38: em_sy.*
39: em_hgto_hum.*
40: em_hgto_mus.*
41: em_hgto_other.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------------------|
| 1 | 2617 | 99.5 | 11421 | 1 | AE001619 Chlamydia |
| 2 | 2617 | 99.5 | 13389 | 1 | AE002202 Chlamydia |
| 3 | 2617 | 99.5 | 110000 | 6 | AR310754_03 Continuation (4 of |
| 4 | 2617 | 99.5 | 110000 | 6 | AR310754_04 Continuation (5 of |
| 5 | 2610 | 99.2 | 300650 | 1 | AP002546 Chlamydia |
| 6 | 2242.5 | 85.3 | 300512 | 1 | AE016995 Chlamydia |
| 7 | 2147 | 81.6 | 14168 | 1 | AE001281 Chlamydia |
| 8 | 2144 | 81.5 | 1587 | 1 | TAJ10586 Chlamydia |
| 9 | 2143 | 81.5 | 23773 | 1 | AE002301 Chlamydia |
| 10 | 1379.5 | 52.5 | 1850 | 6 | AX654307 Sequence |
| 11 | 1379.5 | 52.5 | 1851 | 6 | AX660386 Sequence |
| 12 | 1338.5 | 50.9 | 1872 | 6 | AX654737 Sequence |
| 13 | 1333 | 50.7 | 1875 | 6 | AX505999 Sequence |
| 14 | 1333 | 50.7 | 1875 | 6 | AX651697 Sequence |
| 15 | 1333 | 50.7 | 2110 | 8 | AF428316 Sequence |
| 16 | 1331 | 50.6 | 1933 | 6 | AX660088 Sequence |
| 17 | 1330.5 | 50.6 | 1896 | 8 | AY128844 Arabidops |
| 18 | 1330.5 | 50.6 | 2090 | 8 | AY081350 Arabidops |
| 19 | 1330.5 | 50.6 | 2146 | 8 | AY084374 Arabidops |
| 20 | 1330 | 50.6 | 2181 | 8 | AY045903 Arabidops |
| 21 | 1329 | 50.5 | 2151 | 8 | AY045903 Arabidops |
| 22 | 1306 | 49.7 | 1896 | 6 | AX098423 Sequence |
| 23 | 1306 | 49.7 | 2300 | 8 | STPLADTR Solanum tub |
| 24 | 1302.5 | 49.5 | 2126 | 8 | GSU251356 Citrus hy |
| 25 | 1302 | 49.5 | 2122 | 8 | AY098893 Citrus hy |
| 26 | 1260 | 47.9 | 1770 | 6 | AX098422 Sequence |
| 27 | 1198 | 45.6 | 91001 | 8 | AC013453 Arabidops |
| 28 | 1187 | 45.1 | 132990 | 8 | AC018848 Arabidops |
| 29 | 1157.5 | 44.0 | 2139 | 8 | ATATATP2 Arabidops |
| 30 | 1107 | 42.1 | 174617 | 2 | AP003977 Oryza sat |
| 31 | 1088 | 41.4 | 1503 | 1 | RMO507299 Oryza sat |
| 32 | 1079.5 | 41.0 | 1500 | 1 | CC441310 Caedibact |
| 33 | 1079 | 41.0 | 1497 | 1 | RR1507300 Rickettsi |
| 34 | 1079 | 41.0 | 148018 | 2 | AE003974 Oryza sat |
| 35 | 1076 | 40.9 | 14117 | 1 | AE008575 Rickettsi |
| 36 | 1073 | 40.8 | 1497 | 1 | RTY507301 Rickettsi |
| 37 | 1056 | 40.2 | 1505 | 12 | AF087957 Synchetic |
| 38 | 1056 | 40.2 | 1604 | 1 | R1RATPDDA M28816 R. prowazeki |
| 39 | 1056 | 40.2 | 282610 | 1 | RFX01 Arabidops |
| 40 | 1025.5 | 39.0 | 1823 | 6 | AX100499 Sequence |
| 41 | 1025.5 | 39.0 | 10163 | 1 | AE001646 Chlamydia |
| 42 | 1025.5 | 39.0 | 10407 | 1 | AE002174 Chlamydia |
| 43 | 1025.5 | 39.0 | 110000 | 6 | AR310754_07 Continuation (8 of |
| 44 | 1025.5 | 39.0 | 300550 | 1 | AP002547 Chlamydia |
| 45 | 1020 | 38.8 | 1900 | 1 | AY120885 Holospora |

RESULT 1

ALIGNMENTS

AE001619 11421 bp DNA linear BCT 01-DEC-2000
LOCUS Chlamydia pneumoniae section 35 of 103 of the complete genome.
ACCESSION AE001619
VERSION AE001619.1 GI:4376620
KEYWORDS
SOURCE Chlamydia pneumoniae CML029
ORGANISM Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 11421)
AUTHORS Kalkan, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W.,
Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 99206606
PUBMED 10192388
REFERENCE 2 (bases 1 to 11421)
AUTHORS Kalkan, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
Grimwood, J., Davis, R. W. and Stephens, R. S.
TITLE Direct Submision
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University
of California, 235 Baril Warren Hall, Berkeley, CA 94720, USA
LOCATION/Qualifiers
1. 11421
/organism="Chlamydia pneumoniae CML029"
/mol_type="genomic DNA"
/strain="CML029"
/db_xref="taxon:115713"
/gene="yCgd"
/note="synonym: CPn0346"
complement (95. 1069)
/gene="yCgd"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AAD18490.1"
/db_xref="GI:4376621"
/translation="MALGPSYGVGSFPOFSPFSPRLFGSLFTGSLYINDIOIVT
LAISGAPAGFLVLRKAMANNVSHVTLGLVCVCFTRQTLTSLGTTTLMAMA
TAMLGFILYIRNFVSESSIALVLSLVLVEMTKNAHIGTELVLGNAD
SLTKEDIPVTVILANAVITIFAFRSIVCSFDSVPASLSGIPRLVDYLIIPQLSA
CLVGAFKAVGMALAFILIPSLIAKIVKISRSIMAMSLVSGTAFAPASSRAIL
SAVDLGSISGVFLTMVYIVKFIISVFRGYPSKINPEKISEKSSQ"
complement (1069. 2415)
/gene="yCgc"
/note="synonym: CPn0347"
complement (1069. 2415)
/gene="yCgc"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AAD18491.1"
/db_xref="GI:4376622"
/translation="MLSCVFSDTIFLSSFLAVTLICMTATLWGLTLLISKOPLES
SHASVGLVGLMAOVVSLQASIFVILFEGCASVFGYIIFLGVCKLHMSAL
CPVYVFRALIGTILASVYKESPTLXNRINATLQGAITGELATLAAYFCA
LMMYRQIVITFDKFAVTCGLTALVLSLITSLIVSGVSGVILISAFVA
PSLGRQSDRLSTLLISAFPGISGLAGSYISVAFICRAIIGQAAVPTLPTGPIV
VYCAGLIAGLCLFSPKSGWVIRFRRKIFSKDOEHLKVFHSHNRLENTSVRD
FYCSYKQVDFGPPFRRVQIILEMRGVKKEDYRLTKKGRSEALRLVAHLMW
SYLVNSLPSKSHVLAELIEHVLTELDHLTLTINDPCVDPHRQIIPKKKEV"
complement (2418. 3197)
/gene="yCg8_1"
/note="synonym: CPn0348"
complement (2418. 3197)
/gene="yCg8_1"
/codon_start=1
/transl_table=11
/product="rRNA methylase"
/protein_id="AAD18492.1"
/db_xref="GI:4376623"
/translation="MNKDETFWMSVHNLGVYEHAAVLYHISFLGKSLTALIGPNS

AGKSLTKASLGLIKPSSGTVYFFNOKKRVORLAPORASVDMPFMTVLALAM
GCVYKGMWGRISDDREAFHILERVLESAVDARQGLSGGQOQAFILBALMOKA
DLYIMDELPSAIDMASFKTSYGVLELBDQKTYIVYHHDLSHRQLDHHVLLNKLRL
ICCGPTDECLNGDITIPOTYGCIELEBQTLKSRKQSGC"
complement (3181. 4164)
/gene="yCgA"
/note="synonym: CPn0349"
complement (3181. 4164)
/gene="yCgA"
/codon_start=1
/transl_table=11
/product="Solute Protein Binding Family"
/protein_id="AAD18493.1"
/db_xref="GI:4376624"
/translation="WDAWGYIFKVMRWIFCFVACGTRPGCTSGFONANSRPTLSM
NRMTHDCVERVGNRLATAVLIKSLDPAHEWYKGDOKTAGSAVITCNGLGHEHL
SLRHLLENPNPSVKLERLARGAVPLEEGICDPHIMPLSTWKEVAVITTEFLIE
KFPWSAFKRNCSBELVCEMSILDSMAVPEECSTIBENLRVYSGHNSVYTRRLAT
PEEYASGAFMRSCISPEGLSPEAOISVDIWAVDYIHEHVSVPFEDTINODALKK
IVSSLKSHLVRLAQKPLYSNDVDNVSFTKHNVCLITEELGGVLAECGR"
4575. 5081
/gene="CPn0350"
4575. 5081
/gene="CPn0350"
/codon_start=1
/transl_table=11
/product="CT066 hypothetical protein"
/protein_id="AAD18494.1"
/db_xref="GI:4376625"
/translation="MAVEQSHKEIEIKLGAIKRVGNKENDLCRYLPGPSGYMH
HFTLKMSAPBQLKMLKTFILSESPRTINPKRAPRSGSKRDPINFTKDIER
VELARQGDGDLARFSPKKPLSLKRELIRSRINGIVSELNNAVYEAIVKAVSSPN
LEYSPPV"
5282. 6829
/gene="ad_1"
/note="synonym: CPn0351"
5282. 6829
/gene="ad_1"
/codon_start=1
/transl_table=11
/product="ADP/ATP Translocase"
/protein_id="AAD18495.1"
/db_xref="GI:4376626"
/translation="WMTKEKRFGLKLSFLPHITHELKVLPMFLMFCITFNYTVL
RDTKDLIVAGRSGGAAIPRIKRMVAPCALIMLYAKSNLISAOALFYAATGP
LIFPAIRPTVYIPRDVTHPREPDRLOALIPRELLSGVALLRMTAPRVLAELG
SVMLSLMFWGFANEITKIHEAKRYALFGANISLASGATVWASLGRASVEGD
PMGISLRLLMAMTIVSGLVNASYWMINKVLTDPREYNPEMKSGKGAAPKXNMKD
SFLYLARSPLYLLALVIAVGCINLIEVYWSQLVQYPPNMDSYSEPMGNFSPMTG
VVSVLIMLVFGGVNIRKRGMLTGALVTPMVVLITGIVFALVIRNDSGLVAMEGTT
PLMAVAVGAINONLISKSTKVALPDSIKEMAYIPLDOEKYKGAALDVAARFGSG
GALIQGLVITCGSIGMTPLVAVILLFIAIWLVSATKLKPLAQSALKEQVADE
DSAPASS"
6968. 8227
/gene="CPn0352"
6968. 8227
/gene="CPn0352"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD18496.1"
/db_xref="GI:4376627"
/translation="WKSPLTVPHIFLKNDCGHRSCSKIRPTARLLIGLVAIVS
ALSFVLAAPISIAIGTLAALVILITLVALLKKSKVLPIPNELOKITNRYRK
EIVFYVTHSLTVNELKIFLNCWKSGLDLPNLRKAEAFGIDILKSIDLTFPEFEE
LILQNCPLWYLSHFIDKTESVAGEIGLNKQKVGLGGLPAFHKGYTTFPSSYRPL
TLISESQKFLYSKASKNOWDPSVKTCCEIFRELPHNMIFRDVOGIIQGLFLFS
HGITWEOAKMIGLINPDNWKMLCOFDAQGCSWATRGFLNTEETMDEPVSSNYEPT
VNRTWYKELVLEKYKESPMHRASALVOKICVYTHHQLKMKQVVRNYSQWTS
LPQYAFHQYIKLEKLTSSDPIRSSL"
8287. 8532
/gene="CPn0353"
8287. 8532

QY 461 IIleCyGgISerIleGlyAlaMetThrProTyrLeuAlaValIleLeuLeuPheIleIle 480
Db 6662 ATCTGTGAAGATTGGAGCTATGACCCCTTATCTGAGTGATTCCTTCTTCTACTT 6721
QY 481 AAlaIleTLeuValSerAlaThrLysLeuSulLysLeuPheLeuAlaGlnSerAlaLeu 500
Db 6722 GCTATTGCTGTTGTTCTTGCACCTAAGTTAAACAACCTATTCTTAGCGCAGTCTGCTT 6781
QY 501 LysGIuGIuGIuValAlaGlnGIuLysSerAlaProAlaSerSer 515
Db 6782 AAGAACAAGAGTGTCTCAAGAGATTCACTCTGCTTCTTCA 6826
RESULT 2
AE002202/c 13389 bp DNA linear ECT 30-MAY-2000
LOCUS Chlamydomophila pneumoniae AR39, section 33 of 94 of the complete
DEFINITION genome.
ACCESSION AE002202 AE002161
VERSION AE002202.2 GI:8163425
KEYWORDS
SOURCE Chlamydomophila pneumoniae AR39
ORGANISM Chlamydomophila pneumoniae AR39
REFERENCE 1 (bases 1 to 13389)
AUTHORS White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,
Betty, K., Bass, S., Linner, K., Weidman, J., Khouri, H., Craen, B.,
Bowman, C., Dodson, R., Gwin, M., Nelson, W., Deboy, R., Kolonay, J.,
McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.
Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
pneumoniae AR39
TITLE Nucleic Acids Res. 28 (6), 1397-1406 (2000)
JOURNLT 10684935
MEDLINE 20150255
PUBMED 2 (bases 1 to 13389)
REFERENCE Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F.,
White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,
Betty, K., Bass, S., Linner, K., Weidman, J., Khouri, H., Craen, B.,
Bowman, C., Dodson, R., Gwin, M., Nelson, W., Deboy, R., Kolonay, J.,
McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.
Direct Submission
TITLE Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
JOURNLT Medical Center Dr, Rockville, MD 20850, USA
COMMENT On Jun 1, 2000 this sequence version replaced gi:7189324.
FEATURES
SOURCE Location/Qualifiers
1. .13389
/organism="Chlamydomophila pneumoniae AR39"
/mol_type="Genomic DNA"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
163. .339
/gene="CP0400"
163. .339
/gene="CP0400"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA038245.1"
/db_xref="GI:7189325"
/translation="MLSVSKMLPTQDSIERETDSKDKIIFTYICSSKYLGHFPS
HLDRNKIHESIGV"
597. .1448
/gene="CP0401"
597. .1448
/gene="CP0401"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11

gene
CDS
/product="hypothetical protein"
/protein_id="AA038246.1"
/db_xref="GI:7189326"
/translation="WNIDPVRNTQVNSQATQFSCQALSLIYSVVAATVATV
LVCSQSLSLEIGTALVLSLIFASAMFNIYKRDQPKSLIPKIMELIQEYPSI
VVFISRIQOEVSIVEIHHLISLTKNTNFDAAPYLDKLFQIGKEKDVHPSLQNF
EELIHCPLHMLGRVLVPMVSDVTPEYGYWCGPGLYENAPSLFERSLSLLKXI
SPGFALLEDGLKNTWSSSELVQIRQNLPTRYADKEVDEALNADYQFDSLLHL
IPSHKLS"
1455. .1808
/gene="CP0402"
1455. .1808
/gene="CP0402"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA038247.1"
/db_xref="GI:7189327"
/translation="MQLFOYMSGMDWLCDPDSQGBQSLVGLHSSWALYEA
EQYLPVPSLITWELLEMQLSKPTKGVAKDLNVFEHGFQRFQYLSLDNQAF
ENTFLNPKYHLDR"
1842. .1958
/gene="CP0403"
1842. .1958
/gene="CP0403"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA038248.1"
/db_xref="GI:7189328"
/translation="MKSLSVRSAPFSGIYKSLNLNKKELFMGFFITNS"
2019. .3326
/gene="CP0404"
2019. .3326
/gene="CP0404"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA038249.1"
/db_xref="GI:7189329"
/translation="MAMSWITVPHALFRNHCCHSTPPLSSRTIVRIANSLFCIGL
AALGCLAPVSYVGSVLAFAFVILSLVILALIFBEKLPPTPRIIPDRFTVIDEA
YGLSISAFVREQVTLAEFRQSTALCNLSPEKIKQLPSERSKVESFGISRLAGD
LEKNMNPIDEDLSOTCPDLVWLQKFSAGPOVCRLRPRECYGYWLAGLGYSTKA
ATIRCKETHHIILOQLTKEDVILKKNALDEKMDLVKAIIVERIYTYARGTLKTEA
GGITKRTISKELLLSIHGSPQDOLITQLPDPANDMLCFUNSTRANVQLCALVCA
LSSQNLDSHSDIDFVNLGLVYLQDLKFAVQARSASDEPKELGKFLRLHLSVSKRL
ESVLRQGLHRIALEHGNARARYVDNVFVGARLHRTKTSIFPKO"
complement (3404. .4795)
/gene="CP0405"
complement (3404. .4795)
/gene="CP0405"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA038250.1"
/db_xref="GI:7189330"
/translation="MLNLRNTQIDEPKCMNSITSPVIONNSCNYYELKNSITIH
IVISAILLGCALIAFLCVAAPVSYLISGALLGLIALIGLIGIKITPMISSEKQ
VFPQELVNRBAHYPKFVSPVSEAKNDLISGLIDLNQHSVEGSSNNVASEL
OCKIDMEGIALKKNVETASLKRLESASSRPLPSLAKTIQKVPPFMLGFEISAG
SKVVELHRYAKTIGSLEBLSIDYIKPEMLPTVYLIFLDRPNTSLNLTHTLVAVL
TRDFQHLRYAALNGWNLNHSDLNTHKQDLFKYHAAVQSYHLSQPSIQEDXFTNL
LLCIPFGRYSWKQMSLITVPPADLMENLCLTLDHGRQPDMEFASLIGLYTQGLIH
KESEAFSLSTLSDIQFKTIRQSTNIAMFLENLATNHSSTFRSLPPIIVHPIKRSVF

gene SGPEDSSLLIG
 complement (4776. .5021)
 /gene="CP0406"
 CDS complement (4776. .5021)
 /gene="CP0406"
 /note="hypothetical protein; identified by Glimmer2;
 putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF38251.1"
 /db_xref="GI:7189331"
 /translacion="MRPNIKKSLIFIRIRISQSGKEOKARPEFFKSTISVILL
 LEAFINENESSIIONNPNKNKINSINRIFVKFTI"
 complement (5081. .6361)
 /gene="CP0407"
 CDS complement (5081. .6361)
 /gene="CP0407"
 /note="hypothetical protein; identified by Glimmer2;
 putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF73665.1"
 /db_xref="GI:8163426"
 /translacion="MRHSGCTMKCSPLTVPHIPIKNDCECHRSGLKIRITALLIG
 LVLAVALSIFPLAIPISYALIGTLLAVALITITTVALLAKSKYLPINELQKI
 IYNRPKVEFFVFKTHSLTVNELKFINCMKSGTDLPPNLRKKAFAFIDILKSIDLT
 LEPEBEILIONCPYMLSHFIDKTESVAGEIGLNTKQVGLPLAFHGYTTFH
 SYTRPLTLISSEYKPLYSKASKNOMSPSVKTCSEIPEKELPHNMIFFRDVOGISQ
 FLVFPESGITWEOQMIOLINPDNKKMCCOPDKARHGCMATFEGFNTENMEDPV
 SSVNPEYVNFMTWDELKYLIEKVESPMHPKSAIVQKICVNTTHQNLKRWQFVANT
 SSQRTSSLPQYAFNAQYKLEKKEISSLPINSSL"
 complement (6479. .8026)
 /gene="CP0408"
 CDS complement (6479. .8026)
 /gene="CP0408"
 /note="similar to GB:M28816 SP:P19568 PID:152470
 GB:AJ253269, identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="ADP, ATP carrier protein"
 /protein_id="AAF38252.1"
 /db_xref="GI:7189332"
 /translacion="WTKTBKPRGKLRSEFLPIHTEHAKYLPMPFMTCTTNTTVL
 RDKOTLIVAGPAGSGEALPTKFWLVVPCALPEMLIYAKSNISKOALFYAGTPE
 LIFPALPPTVIVPRDVAHPEDRLQALPPGLGLVALIRNMTFPAFVLELWG
 SVMLSIMFGFANEITKIHAKRFYALFGIGANISLASGRALVASKLRAVSSEGVD
 PMGISLRLLMAMTIVSGILMASVYMKNVLTDPFPYNPPEMOKGKKGAKNNMKD
 SEIVYARSPYITLLALIVYAGICINLEWTKSOLKQYPRANDYSFPMNPFMTG
 VSVYVILFVGSNVIRKFGMLGALVTFVMDLTGIVFPAIVIFRNQSGVLMFMTG
 PLMAVAVGAIONISKSTKVALFDSTEMVYIPUDQKXVGKKAIDIVARFEGSG
 GALIOGGLVIGSIGAMTPYLAIVILFIIAIMVLSATKMLFLAQSALKEOEVAOE
 DSAPASS"
 complement (8227. .8742)
 /gene="CP0409"
 CDS complement (8227. .8742)
 /gene="CP0409"
 /note="conserved hypothetical protein; identified by
 Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF38253.1"
 /db_xref="GI:7189333"
 /translacion="MGFMAVEQSHIKEIEKLGKAIKRVCGKNENDLCRYLPGPSG
 YMHFTLKMKSSAAPEOLLKMLKTFILSEPRPTINPBRAPRGSKRRDPINFKTD
 IERVLEARQNDKDLARFSPKRLTSLKELISINGLIVSVMNAYVBAKAVAS
 SPNLEVTSPV"

Alignment Scores:
 Pred. No.: 6,736-222 Length: 13389
 Score: 2617.00 Matches: 513

| Query | Percent Similarity: | Best Local Similarity: | Query Match: | Conservative: | Mismatches: | Indels: | Gaps: |
|--|---------------------|------------------------|--------------|---------------|-------------|---------|-------|
| US-09-869-433-2 (1-515) x AB002202 (1-13389) | 99.61% | 99.61% | 99.51% | 0 | 2 | 0 | 0 |
| 1 MetThrylsThrguJugluveProPheGlylysLeuArgSerPheLeuTrpProIleHis 20 | | | | | | | |
| 8026 ATGACAAAACCGAAGAAAACCTTTGGAAAATTGGCCTTTCTTGTGGCGATACAT 7967 | | | | | | | |
| 21 ThrHisgluLeuYslYsValleuProMetPheLeuMetPhePheCysIleThrPheAsn 40 | | | | | | | |
| 7966 ACTACAGCATTAAGAAAGTCTGACATGTTCCTAAGTTCTTGTGATTAACATTTAAC 7907 | | | | | | | |
| 41 TyrThrValleuArgAspThrLysAspThrLeuIleValGlyAlaProGlySerGlyVala 60 | | | | | | | |
| 7906 TATACGGGTTCACCGATACAAAGACACTCTTATTGTGGAGGCTCCGTGTTCTGTGCA 7847 | | | | | | | |
| 61 GluAlaIleProPheIleLysPheThrLeuValValProCysAlaIleIlePheMetLeu 80 | | | | | | | |
| 7846 GAGGCAATACCTTCATCAAGATTGGCTGTGTGCTCCCTGTGCTATTATCTTATGCTT 7787 | | | | | | | |
| 81 IleTyrAlaYslLeuSerAsnIleLeuSerIysGluAlaLeuPheYrAlaValGlyThr 100 | | | | | | | |
| 7786 ATTTATGCAAGCTTAAGTAATTTTAAGTAAGCAGGCTTATTTATGCAAGTGGAAACG 7727 | | | | | | | |
| 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValleu 120 | | | | | | | |
| 7726 CCTTTTAAATTTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7667 | | | | | | | |
| 121 HisProThrGluPheAlaAspArgLeuAlaIleLeuProProGlyLeuLeuGlyLeu 140 | | | | | | | |
| 7666 CATCTTACAGAAATTGCTGACCGCTTACAGGCACTCACTCACTCACTCACTCACTCACT 7607 | | | | | | | |
| 141 ValAlaIleLeuArgAsnTrpThrPheAlaIlePheYrValleuAlaGluLeuTrpGly 160 | | | | | | | |
| 7606 GTTGCCATCTTAAGAACTGACATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7547 | | | | | | | |
| 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleHisGlu 180 | | | | | | | |
| 7546 AGCGTCATGCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7487 | | | | | | | |
| 181 AlaIysArgPheYrAlaLeuPheGlyIleGlyValleuIleSerLeuLeuAlaSerGly 200 | | | | | | | |
| 7486 GCAAGGCTTTCTACGCTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7427 | | | | | | | |
| 201 ArgAlaIleValTrpAlaSerLysLeuArgAlaSerValSerGluGlyValAspProTrp 220 | | | | | | | |
| 7426 CTTGCAATGTTTGGGCTTCAAGATTGAGACCTTCCGTTTCTGAAGGTGATGATCTTGG 7367 | | | | | | | |
| 221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValleuMetAla 240 | | | | | | | |
| 7366 GGAATTTCTTACGCTCTTTGATGCTATGATGCTATGATGCTATGATGCTATGATGCTAT 7307 | | | | | | | |
| 241 SerTyrTrpTrpIleAsnLysAsnValleuThrAspProAlaGlyPheYrAsnProGlu 260 | | | | | | | |
| 7306 AGTTACGTGTGATCAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 7247 | | | | | | | |
| 261 MetGluYslGlyLysGlyAlaLysProLysMetAsnMetLysAspSerPheLeuTrp 280 | | | | | | | |
| 7246 ATGCAAAAGGGAAAAAGTGTCTTAACCTTAATTAATTAATTAATTAATTAATTAATTAAT 7187 | | | | | | | |
| 281 LeuAspArgSerProYrIleLeuLeuLeuThrLeuLeuValIleAlaYrGlyIleCys 300 | | | | | | | |
| 7186 CTTCGTGATCTCTCTTATATCTTTTATAGCTCTCTGTTATGCTTATGCTTATGCTTATGCT 7127 | | | | | | | |
| 301 IleAsnLeuIleGluValThrTrpLysSerGluLeuYslLeuGluIleYrProAsnMetCasn 320 | | | | | | | |
| 7126 ATTAACCTTAATGAGAGTCACTGAAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 7067 | | | | | | | |
| 321 AspTyrSerGluPheMetGlyAsnPheSerPheTrpTrpGlyValIleValSerValleuIle 340 | | | | | | | |

Db 7066 GACCTATAGAGACTTACATGGGGAACCTCTCCTTCGTGACCTGGCGGTAGATCCGACTATATC 7007
 Qy MetLeuPheValIGIYGIYsEnValIIeaGLYsPheGIYTrPLeuThrgIYAlaLeuVal 360
 Db ATGCTATTATTGGTGGTGGTAACGCTCATTCGTAATTGGATGGTTAACTGGAGCCCTAGTC 6947
 Qy ThrProValMetValLeuLeuThrgIYIleValPhePheAlaLeuValIIePheArgAsn 380
 Db ACTCCTGTCATGATGTTCTCTTAACAGGATACGTTTCTTGCTCTTGTATCTTTAGAAC 6887
 Qy GlnAlaSerGIYLeuValAlaMetPheGIYThrThrProLeuMetLeuAlaValVal 400
 Db CAAGCTTCGGGCTGGCTGGCTATGTTGGGTAACACTCCTCTCATGTAGCTGTGGTTGTC 6827
 Qy GIYAlaIIeGLIAsnIIeLeuSerIYsSerThrIYsTrAlaLeuPheAspSerThrIYs 420
 Db GGAGCTATACAGAAATATCTTCTTGAATACACAAATACGCTCTCTTGTACTCACTAA 6767
 Qy GIIuMetAlaTYrIIeProLeuAspGInGInIYsValIYsAlaIIeAsp 440
 Db GAAATGGCCTATATATCCTCTTGACCAAGACAAAGTCAAGTAAAGGCTGCTATTGAT 6707
 Qy ValValAlaAlaArgPheGIYIYsSerGIYAlaLeuIIeGInGInIYsLeuVal 460
 Db GTAGTTCGCCGCCCTCGCTGGAAATACAGAGAGCTTTAATCAACAGTTTGCCTGTT 6647
 Qy IIeCYsGIYserIIeGIYAlaMetThrProTYrLeuAlaValIIeLeuPheIIe 480
 Db ATCTGTGGAAGTATGGAGCTATGACCCCTTATCTTGACGTGATCTTCTTTCATCAT 6587
 Qy AlaIIeTrPLeuValSerIaThrIYsLeuAsnIYsLeuPheLeuIIaGInSerAlaLeu 500
 Db GCTATTGTGGTGGTTCGCACTAAGTTAAACATACTTCTTACGCGAGTGTGCTCTT 6527
 Qy 501 IYsGIuGInGInIYsValAlaGInGInIYsSerAlaProAlaSerSer 515
 Db 6526 AAAGAACAAAGAGTGGCTCAAGAGATTCAGCTCCCTGCTTCTTCA 6482

RESULT 3
 AR310754_03
 WPCOMMENT
 Sequence split into 13 fragments LOCUS AR310754 Accession AR310754

| Fragment Name | Begin | End |
|---------------|---------|---------|
| AR310754_00 | 1 | 110000 |
| AR310754_01 | 100001 | 210000 |
| AR310754_02 | 200001 | 310000 |
| AR310754_03 | 300001 | 410000 |
| AR310754_04 | 400001 | 510000 |
| AR310754_05 | 500001 | 610000 |
| AR310754_06 | 600001 | 710000 |
| AR310754_07 | 700001 | 810000 |
| AR310754_08 | 800001 | 910000 |
| AR310754_09 | 900001 | 1010000 |
| AR310754_10 | 1000001 | 1110000 |
| AR310754_11 | 1100001 | 1210000 |
| AR310754_12 | 1200001 | 1230025 |

Continuation (4 of 13) of AR310754 from base 300001 (AR310754 Sequence 1 from patent US

Alignment Scores:
 Pred. No.: 9,71e-221 Length: 110000
 Score: 2617.00 Matches: 513
 Percent Similarity: 99.61% Conservative: 0
 Best Local Similarity: 99.61% Mismatches: 2
 Query Match: 99.51% Indels: 0
 DB: 6 Gaps: 0

US-09-869-433-2 (1-515) x AR310754_03 (1-110000)

Qy 1 MetThrIYsTrhGInGInIYsProPheGIYIYsLeuArgSerPheLeuTrPProIleHis 20
 Db 104370 ATGCAAAACCAAGCAAAACCTTTTGGAAAATTCGCTCTTTCTTGTGGCCGATACAT 104429
 Qy 21 ThrHisGInLeuIYsIYsValLeuProMetPheLeuMetPhePheCysIIeThrPheAsn 40

| | | | |
|----|--------|---|--------|
| Db | 104430 | ACTCAGCAGCTTAAGAAAGTTCTGCGAATGTTCTCTTAATGTTCTTCTGATTAACATTTAAC | 104489 |
| OY | 41 | TyrThrValLeuArgAspThrLysAspThrIleuIleValAlaProGlySerGlyAla | 60 |
| Db | 104490 | TATACGGGTGTACCGGATACAAACACCTCTTATGTGGGAGCTCCCTGGTCTGGTGA | 104549 |
| OY | 61 | GluAlaIleProPheIleLysPheTrpIleuValProCysAlaIleIlePheMetLeu | 80 |
| Db | 104550 | GAGCGAAATACCTTTCATCAAGTTTGGCTGTGTGCCCTGGCTATTAATCTTAATAGCTT | 104609 |
| OY | 81 | IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr | 100 |
| Db | 104610 | ATTATATGAAAGCTTAAGTAATATTTTAAGTAAGCAGCGCTTATTTATCAGTGGGAAG | 104669 |
| OY | 101 | ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu | 120 |
| Db | 104670 | CCCTTTTAAATTTCTTCTTGCCCTGTGCCACTGTAATTAATCCGCTAAGCATGTTTAA | 104729 |
| OY | 121 | HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu | 140 |
| Db | 104730 | CATCCCTACAGAAATTTGCTGACCGTTTAAAGGCATCTCACTCCAGAAATTTGCTAGGAC | 104789 |
| OY | 141 | ValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTyrValLeuAlaGluLeuTrpGly | 160 |
| Db | 104790 | GTTGCCATCTTAAGAAACTGACATTTCTGCAATTTATGTACTTGCTGAACATAGGGA | 104849 |
| OY | 161 | SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleHisGlu | 180 |
| Db | 104850 | AGCGCATGATCTCTCTAATGTTCTGGGGAATTTGCTAATGAATTAACAAATAATCCACGA | 104909 |
| OY | 181 | AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly | 200 |
| Db | 104910 | GCAAGCGCTTCTACGCTCTTTCGGTACGAGCTAATATTTCTTAATCACTTCTGAT | 104969 |
| OY | 201 | ArgAlaIleValIleTyrAlaSerLysLeuArgAlaSerValSerGluGlyValAspProTrp | 220 |
| Db | 104970 | CGTGCAATGTTTGGGCTTCAAAGTGGAGCTTCCGTTTTCGAAGGTCTAATCCTTGG | 105029 |
| OY | 221 | GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla | 240 |
| Db | 105030 | GGAATTTCTTAACGCTTTTGTATGGCTATGACTATGTATGCTGACCTGTTCTTAATGGCC | 105089 |
| OY | 241 | SerTyrTrpTrpIleAsnLysAsnValLeuThrAspProArgPheTyrAsnProGluGlu | 260 |
| Db | 105090 | AGTACTGGTGGATCAATTAACAAAGTATGACCGATCTCTCGCTTAAATCAACAAAGA | 105149 |
| OY | 261 | MetGlnLysGlyLysLysGlyAlaLysProLysMetAsnMetLysAspSerPheLeuTyr | 280 |
| Db | 105150 | ATGCAAAAAGGGGAAAAAGGCTTAACCTTAATGAATATGAAGAATAGTCTTCTTAT | 105209 |
| OY | 281 | LeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCys | 300 |
| Db | 105210 | CTTGCTAGATCTCTTAATATTTCTTTATTAAGTCTCTGTTATATGGCATATGATTTG | 105269 |
| OY | 301 | IleAsnLeuIleGluValThrTyrLysSerGlnLeuLysLeuGlnTyrProAsnMetCys | 320 |
| Db | 105270 | ATTAACTTAATCGAAGTGAAGTGAAGAAAGTCAAGTGAACATCGCAATATCTTAATGAAT | 105329 |
| OY | 321 | AspTyrSerGluPheMetGlyAsnPheSerPheTrpTrpGlyValValSerValLeuIle | 340 |
| Db | 105330 | GACTATAGTGGTTCATAGGGGAACCTTCTCTTGAGCTGGGGTATGATCCGTAATATC | 105389 |
| OY | 341 | MetLeuPheValGlyGlyAsnValIleArgLysPheGlyTrpLeuThrGlyValAlaLeuVal | 360 |
| Db | 105390 | ATGCTATTTTGTGGTGTGAAGCTATCTGTAATTTTGAATGGTTAACTGGAGCCCTAGTC | 105449 |
| OY | 361 | ThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn | 380 |
| Db | 105450 | ACTCTGTGCATGGTTCCTCAACAGATATCGTTTCTTCCGCTGTGATCTTTTGAAGAC | 105509 |
| OY | 381 | GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValValVal | 400 |

| | | | |
|----|------|--|------|
| Db | 4610 | ATTATATGAAAGCTAAGTAAATATTTTAAGTAAAGAGCCCTTATTTATGCACTGGAAAC | 4665 |
| QY | 101 | ProPheLeuIlePhePheIleAeuPheProThrValIleTyPProLeuAArgAspValIleu | 120 |
| Db | 4670 | CCCTTTTAATTTCTTTGCGCTGTGCCGACTGTAATTTATCCGCTACGGCACTTTTAA | 4722 |
| QY | 121 | HisProThrGluPheIleAspArgLeuGlnAlaIleLeuProProGlyIleuLeuGlyIleu | 140 |
| Db | 4730 | CATCCCTACAGAAATTTGCTGACCGCTTACAGGCCATCCCTACCTCCAGGATTTGCTAGAGCTC | 4783 |
| QY | 141 | ValAlaIleLeuAArgAsnTrpThnPheAlaIaPheTyPValIleuAlaGluLeuTrpGly | 160 |
| Db | 4790 | GTTGCGATCTTAAGAACTGGAACATTGCTGCATTTTAATGATCTTGCAACTAAGGGGA | 4844 |
| QY | 161 | SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleHisGlu | 180 |
| Db | 4850 | AGCGTCATGCAATCTCTTAATGTTCTGGGGAATTTGCTAATGAATAACAAAATCCACAGA | 4900 |
| QY | 181 | AlaIysArgPheTyPAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly | 200 |
| Db | 4910 | GCAAAAGCTTTCTACGCTCTTTGCGATCCGAGCTAATATTTCTTACTAGCTTCGTGT | 4965 |
| QY | 201 | ArgAlaIleValITrpAlaSerLysLeuAArgAlaSerValSerGluGlyValAspProTrp | 220 |
| Db | 4970 | CGTGCATATGTTGGGCTTCAAAAGTGGAGACCTCCGTTCTGAAAGGTGTGATCTTTGG | 5023 |
| QY | 221 | GlyIleSerLeuAArgLeuLeuMetAlaMetThrIleValISerGlyIleValLeuMetAla | 240 |
| Db | 5030 | GGAATTTCTTTACGCTTTTGAATGGCTATGACTATGATCTGACACTTGTCTTATGGCC | 5085 |
| QY | 241 | SerTyPTrpTrpIleAsnLysAsnValIleuThrAspProArgPheTyPAsnProGluGlu | 260 |
| Db | 5090 | AGTTACTGCTGGATCAATTAAGAAAGTATGACCCGATCTCCGCTCTTAATTAACAAAGAA | 5145 |
| QY | 261 | MetGlnLysGlyLysLysGlyValAlaLysProLysMetLysAsnSerPheLeuTyP | 280 |
| Db | 5150 | ATGCAAAAGGGGAAAAAAGGTGCTAAACCTAAATGAATATGAAGAATAGCTTCTCTAT | 5203 |
| QY | 281 | LeuAspArgSerProTyPIleLeuLeuLeuThrLeuLeuValIleAlaTyGlyIleCys | 300 |
| Db | 5210 | CTTGCTGATGATCTCTTAATATCTTTATTAAGCTCTCTTGATTTAGCTTAAGATTTGTC | 5265 |
| QY | 301 | IleAsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGlnTyPProAsnMetAsn | 320 |
| Db | 5270 | ATTAACTTAATCGAAGTACCTTGGAAAAAGTCAGCTGAACGCAATATCCTAATATGAT | 5322 |
| QY | 321 | AspTyPArgGluPheMetGlyAsnPheSerPheTrpThrGlyValValSerValLeuIle | 340 |
| Db | 5330 | GACTATATGATGATTCATGCGGAACTTCTCTTCGAGCTGCGGTGATCCGACTCTTATC | 5383 |
| QY | 341 | MetLeuPheValIGlyIleAsnValIleArgLysPheGlyTrpLeuThrGlyValaLeuVal | 360 |
| Db | 5390 | ATGCTATTATTGTGGTGAACGCTATTCGTAATTTGATGGATGTTACTGAGCCCTAGTC | 5443 |
| QY | 361 | ThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn | 380 |
| Db | 5450 | ACTCCTGTCATGCTTCTCCTTAACAGATATGCTTTTCTTCGCTCTTGTATCTTTAGAAGAC | 5500 |
| QY | 381 | GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValValVal | 400 |
| Db | 5510 | CAAGCTTCTGGGCTGGTGGCTATGTCGGTCAAACTCCTTCATGCTACGTGCTGTGTC | 5565 |
| QY | 401 | GlyAlaIleGlnAnuIleLeuSerLysSerThrLysTyPAlaLeuPheAsnSerThrLys | 420 |
| Db | 5570 | GGAGCTATACGAATATCTTTCGAAATCCCAAAATAGCTCTCTTGACTCAACTAAA | 5623 |
| QY | 421 | GluMetAlaTyPIleProLeuAspGlnGluGlnLysValIleLysGlyLysAlaIleAsp | 440 |
| Db | 5630 | GAAATGGCTATATCCCTCTTGACCAAGCAAGAAAAGTCAAAAGTAAAGCTGCTATTTGAT | 5688 |
| QY | 441 | ValValAlaIaArgPheGlyLysSerGlyGlyAlaLeuIleGlnGlnGlyLeuLeuVal | 460 |

Db 5690 GTAGTTCGCCGCCGCTTCGAAATCAGAGAGCTTTAATCAACAAGATTTCCTT 5749

Qy 461 TleCysGlySerIleGlyAlaMetThrProTyrIleuAlaValIleuLeuPheIleIle 480

Db 5750 ATCTGTGAAGATATTGGAGCTATGACCCCTTATCTTGAGAGATTCTTTTATCATCTT 5809

Qy 481 AAlaIleTPluValSerAlaThrLysLeuLeuLysLeuPheLeuAlaGlySerAlaLeu 500

Db 5810 GCTATTGTGTTGTTCTTCGCAACTAAGTTAAACAACATTTCTTATGCGCAGTCTCTCTT 5869

Qy 501 LysGluGluGluValAlaGluGluAspSerAlaProAlaSerSer 515

Db 5870 AAAGAACAGAGAGTGGCTCAAGAGATTCACTCTCTCTTCA 5914

RESULT 5

AP002546 300650 bp DNA linear BCT 25-MAY-2002

LOCUS Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,

DEFINITION section 2/4.

ACCESSION AP002546 AB033782 AB033783 AB033784 AB033785 AB033800 AB033801

AB033802 AB033803 AB033804 AB033805 AB033806 AB033807 AB033808

AB033809 AB033810 AB033811 AB033812 AB033813 AB033814 AB033815

AB036079 AB036080 AB036081 AB036082 AB038348 AB038349 BA000008

AP002546.2 GI:10176692

VERSION Chlamydomophila pneumoniae J138

KEYWORDS Chlamydomophila pneumoniae J138

SOURCE Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.

ORGANISM Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.

REFERENCE 1

AUTHORS Shirai,M., Hirakawa,H., Ouchi,K., Tabuchi,M., Kishi,F., Kimoto,M.,

Tateuchi,A., Nishida,J., Shibata,K., Fujinaga,R., Yoneda,H.,

Matsushina,H., Tanaka,C., Furukawa,S., Miura,K., Nakazawa,A.,

Ishii,K., Shiba,T., Hattori,M., Kuhara,S. and Nakazawa,T.

TITLE Comparison of outer membrane protein genes omp and pmp in the whole

genome sequences of Chlamydia pneumoniae isolates from Japan and

the United States

J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)

20298986

PUBMED 10839753

JOURNAL

MEDLINE

REFERENCE 2

AUTHORS Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,

Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.

TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138

from Japan and CML029 from USA

Nucleic Acids Res. 28 (12), 2311-2314 (2000)

20330349

PUBMED

REFERENCE 3

AUTHORS Shirai,M.

TITLE Direct Submission

SUBMITTED (04-JUL-2000) Mitsunori Shirai, Yamaguchi University

School of Medicine, Department of Microbiology, 1-1-1

Mitsunori Shirai, Ube, Yamaguchi 755-8505, Japan

(E-mail:mshirai@po.cc.yamaguchi-u.ac.jp, Tel:81-836-22-2227,

Fax:81-836-22-2415)

COMMENT On or before Sep 15, 2000 this sequence version replaced

gi:61722290, gi:61722292, gi:61722294, gi:61722296, gi:61722326,

gi:6172338, gi:6172330, gi:6172332, gi:6172334, gi:6172336,

gi:6172338, gi:6172382, gi:6172384, gi:6172386, gi:6172388,

gi:6172390, gi:6172392, gi:6172394, gi:6172396, gi:6174668,

gi:6655174, gi:6655176, gi:6655178, gi:6655180, gi:8547433,

gi:8547438, gi:8978640

FEATURES

SOURCE

1. .300650

/organism="Chlamydomophila pneumoniae J138"

/mol_type="Genomic DNA"

/strain="J138"

/db_xref="taxon:138677"

/note="synonym:Chlamydia pneumoniae (strain J138)"

144. .935

/gene="CpJ0267"

144. .935

/gene="CpJ0267"

gene

CDS

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BA08477.1"

/db_xref="GI:8978641"

/transl_table=11

/translation="MSIMSLINKTNALINQEPAPVCUNAMPKRYINDRKPACTVTL

VITLMLTIGTVILVILAMSGPGLSVSTIGTSVTLTALFIVGLIKSLAM

OYKVEVVKORKEPSPRIKPNNDVHKLSTCLSPDLIESPSPASTPVSKRIACSG

YALVGLTLLIGAVSVIFPCTGYQLATCTGFCIGTALFVGGIAGARTHSILAQIM

YLVTLYLSALSERNEVTKDORNEINVTYLTREOROKREKALLE"

957. .1343

/gene="CpJ0268"

957. .1343

/gene="CpJ0268"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BA098478.1"

/db_xref="GI:8978642"

/transl_table=11

/translation="MSQCSSSTSTWEMKSPVPMNKNPPTSPSPSEDEFILAYEP

FVLPKTPDPAKQANPGETSPVYENGIDLPILGQNEQNANNPOTSSNPTSLPA

PERLPETBENSQSEBQSQNNBDLIG"

complement (1495. .2472)

/gene="CpJ0269"

complement (1495. .2472)

/codon_start=1

/transl_table=11

/product="dipeptidase"

/protein_id="BA098479.1"

/db_xref="GI:8978643"

/transl_table=11

/translation="MTINDLCSHPHPCRKDPAYRCSPPELLSGSVROVCAIFP

HSGERPCDKQNSLFSLSNPQYDGLSLSTREBENGSSSKSLSTRISMAALD

DVPLGLTLKLTILTRQGPALAYLQWKNDNFGGTGTEKRLSNDKVLIMYEL

GVPLIDSHCSGDAEDILDYADKLPLAVIASHSNRSVLDPRNLDVDAHEVIR

KGVIGLNLVRSYVGDLSGLDEKHLAEINGLISVSLADPEFPAENDEPFPNECS

AEAPVNLQILHRIFFSKAKESILSSRAEKFLLQVIVEQVNPKITDKL"

complement (2487. .3347)

/gene="yw1C"

complement (2487. .3347)

/codon_start=1

/transl_table=11

/product="Sua5 superfamily-related protein"

/protein_id="BA098480.1"

/db_xref="GI:8978644"

/transl_table=11

/translation="MPDKKAQITFSLPVNSAIHQKIVALPDTVYGFVLSLYASEA

EEELVALKOREPSKAPALVYNSIEDENISGYLSPTAKLAQLFPGAILTVKHNRP

RFPKETLAFPIVDHSVREIVDHGCTLIGTSANLSEFPAALTAQEIFADPADLCLIF

DGPCSHGDESTVAVSDPIVYREGILSRVTEINAGTEAKIFRSTHAFSKHILKTV

KNOEOVSPISGSLDKGVCEHPKRNPFYTRLEALKKKTPSIVFIYDINTSDYBEL

FPFLSPFYIE"

3656. .4387

/gene="CpJ0271"

3656. .4387

/gene="CpJ0271"

/codon_start=1

/transl_table=11

/product="lysophospholipase esterase"

/protein_id="BA098481.1"

/db_xref="GI:8978645"

/transl_table=11

/translation="MTDYSFPRKIKNIEAIECRGNPDPIIILCHGYSLADNLTFP

PSICFSKLRPTIIPNGIILPNDPRGNSACPLAVLILDELSTRILYANGVGLQERY

DELFDVLETPKALELILNLNRPYNEIILIGFSQALILHLVLSQNPVAGLIF

AGARLNVQKMEGLKQCAQVPLQSHGDEDEILPYHLGAHLNDLLTLKNGQVSPFG

GHEIPSVFQKQVTVPMNIDPARG"

complement (4359. .5255)

/gene="dnaX_2"

complement (4359. .5255)

/codon_start=1

/transl_table=11

/product="DNA polymerase III gamma and tau"

gene

CDS


```
/protein_id="BA98482.1"
/db_xref="GI:8978646"
/translat="MHEEENOGWEALLRVYHOEVPATILHGETLPVQDRAEOLA
SEILSSPSSSEHKVSOKTHPDIOFPEEGKRLHSIDLPRGKIQIYISFEANTKI
YIIHEADRTMLAAISAFKVEEPPKIAVILITPAKQVLPKTIISRLSIFLERGK
ILCSGSESYLFRPAOCEIPVTEVSQIIXSSSEKQVLDKQRFMEVLEEDRDY
TLNLGLASALANYEHEVKIILQLPLDLKVLILVESACSLNNSSSAASVLEVAIO
LVSLQYKEKELVSPPGDLN"
complement (5246. .5866)
/gene="vdk"
complement (5246. .5866)
/codon_start=1
/transl_table=1
/product="thymidylate kinase"
/protein_id="BA98483.1"
/db_xref="GI:8978647"
/translat="MPVIEGEGSGSKSLAKALGQVADRKVLITREPGLIGE
RLRDLLEPHELESCCELFLPGSRAGHIOEYIIPALDGYVICERHDSITVYQ
GIAEGADPVDLCSKVGPPELPPLPVLDDIPADIGLGRKHROKFPKPKPLS
YHNRIRBGFLSIASADPSRYLVDARSLSALDKWHLHTQGLCT"
complement (5871. .8375)
/gene="gyra.1"
complement (5871. .8375)
/gene="gyra.1"
complement (5871. .8375)
/codon_start=1
/transl_table=1
/product="DNA gyrase subunit A"
/protein_id="BA98484.1"
/db_xref="GI:8978648"
/translat="MFKKDEILVKNLEEMKESYLRSYMSVILSRALPDIRDLKLS
ORVLYAMKOLSLSPGAKHRCCKAIGDTSGDYHPHESISYPLVPMANMRYPL
VDOGNFGSIDDBPRAMRYTERALHTSANTLMEDDKQVVDIVPYNDETRKHPVP
SKPNNLCNGSSGANGMATNIPPHNIGELIEATLLALPQSVSDILQVMPGDPD
TGGIICGSEGIIRSTYTGKIKIRAVLAHEENEDKRESIIITEMFYNNKRLLEQ
IAMLVNEKTLAGISDVDESDDKGIIVLEIKKGESEIILNLYKFTDQVTEGAM
LADLKNLPTMSIHRMISAMIRHREKIVIRRTREYELKAEATRAVLEGYKALSCDA
LVKTIRESGKHEAKERIIRSGFTEPOALAIETRLVQYGLAEKIOKEVEELNK
IAYKQVLSDEGLVKDIIKRELQDLKHHVARTTIEFPADDIRDIEDIITNESVI
TISGDDYVKMPVAVKEQRGGHGVTFPMKGAFLKAVSAFTQVLLITFNEQ
CYWLKQWQPEGGERAKKPIINLEBGRIGEBELALANTKNFDMAGFLFLATKRGV
KAYLSAFLNPRKKIGRILEIDEBELIACHIVSDEKYMFLHLMVAVFPHEKVR
PMGRTAGVGVGLKNEDEKVSQCIIVTEROSVLIQDQFGKSLVEDPRTNRGV
GVSLIENENGVLGAIPIVTDHDSIILMSOGQAIRINODVAVKRSIQGRVLAHL
KEEDALVSMKSSNENDDEVLSGSEECSDTVSLR"
complement (8391. .10808)
/gene="gyrB.1"
complement (8391. .10808)
/gene="gyrB.1"
complement (8391. .10808)
/gene="gyrB.1"
complement (8391. .10808)
/codon_start=1
/transl_table=1
/product="DNA gyrase subunit B"
/protein_id="BA98485.1"
/db_xref="GI:8978649"
/translat="MDPKKQVDAATVLEGLQAVBERPQMTIGDTGTLHLHYVE
VUNSIDEMAGYCSRIDVRLIDEGIVITDNGRGIPIEVBESAKQSGEVALYVE
LITLHAGKFDKDSYKSGGLHGVSCVNAISEKVATYFKKCTOMFSPGCIPT
PLQVSVSDQGEIVFYDPKIFSTCTPDRSILMKRLALNKRITTVLEDDRV
SFPKVFYEGGIQSFVSYLNQNKESLFPSPYICGRVDDDEIEFEALQNNNGS
ELVSYVANNIPTRQGTHTLGTSTALRVNNTYIKANLAKNNKLATGIDREGVLA
VISVKNVPOFEGOTKOLGNSDVSVAOVGVALIIPRENDOLARMIVDVAVPA
QABEAKKRELTRKSLADLSARLPGILDCLEPKGCMYVEGDSAGSAGKQGD
RRFOALIPRIGKILNVEKARLOKIPONOEIGTIIAALGCGTADNFSLKLRITII
MTDADVDSGIRITLLTFYRHTMALIENCCVYIAOPLVSKKQDFRILISEKED
SYLMGLTNESSILFKSTERELGEALLESINVLDESFINLLEKALIPFSEFLEY
KEGIGYLVLAIPATGMOGRVYVSDKEEALAOBETHFKIIELVKAVPVDIQW
LKEGGLDISYLLPOKNEIYIENEDSPSCVSCYTLLEVINYLNIGRKGIEIORYKG
LGEMLADMLDITMNPQRILIHVSLKDAVBADHITFMNGEVPFRPRETIESHALSI
RINNDI"
complement (10812. .11159)
gene
1.45e-219 Length: 300650
```

```
Score: 2610.00 Matches: 512
Percent Similarity: 99.42% Conservative: 0
Best Local Similarity: 99.42% Mismatches: 3
Query Match: 99.24% Indels: 0
DB: 1 Gaps: 0
US-09-869-433-2 (1-515) x AP002546 (1-300650)
Qy 1 MetThrLysThrgLugLysPProPheGlyLysLeuAArgSerPheLeuTTPProIleHis 20
Db 93996 ATGACAAAACCGAAGAAAACCTTTGGAAAATGGCTCTTCTTGTGGCCGATCAT 94055
Qy 21 ThrHisGluLeuLysLysValLeuProMetPheLeuMetPhePheCysIleThrPheAsn 40
Db 94056 ACTCAGAGCTTAAGAAAGAGTTCTGCCAATGTTCTTAATGTTCTTGTATTAACCTTTAAC 94115
Qy 41 TyrThrValLeuArgAspThrLysAspThrLeuIleValGlyAlaProGlySerGlyAla 60
Db 94116 TATACGGGTGTTACCGCATCAAAAAGACACTTATTGTGGAGCTCTGGTTCGTGCA 94175
Qy 61 GluAlaIleProPheIleLysPheThrPheValAlaProCysAlaIleIlePheMetLeu 80
Db 94176 GAGGCAATACCTTTCATCAAGTTTGGCTTGTGTCCCTGTGTATTAATCTTATGCTT 94235
Qy 81 IleTyrAlaLysSerSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr 100
Db 94236 ATTTATGCAAGCTTAAGTAAATATTTTATGAGAGAGCTTATTTTATGCAATGGAGAACG 94295
Qy 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
Db 94296 CCGTTTATATTTTCTTGTGCGCTTCCGACGTGATTTATTCGATACGAGATGTTTAA 94355
Qy 121 HisProThrGluPheAlaAspAlaGluGlnAlaIleLeuProProGlyLeuGlyLeu 140
Db 94356 CATCCCAAGAAATTTGCTGACCTTTTACAGGCACTCTTACCTCAGATTTGCTAGCTC 94415
Qy 141 ValAlaIleLeuArgAsnTPTrPheAlaAlaPheTyrValLeuAlaGluLeuTrrGly 160
Db 94416 GATGCCATCTTAAGAAACCTGAGACATTTGCTGCAATTTATTTACTTGTGAACTTAGGGA 94475
Qy 161 SerValMetLeuSerLeuMetPheTrrGlyPheAlaAsnGluIleThrLysIleHisGlu 180
Db 94476 AGGTCATGCTATCTCTAAGCTTCTGGGATTTGCTAATGAATTAACAAAATACAGAA 94535
Qy 181 AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
Db 94536 GCAAGCGTTTCTACCGCTCTTTCGATCGAGCTAATATTTCTTACCTAGCTTCTGCT 94595
Qy 201 ArgAlaIleValIrrPAlaSerLysLeuArgAlaSerValSerGluGlyValaAspProTrr 220
Db 94596 CGTGCAATGTTTGGGCTTCAAGGTTGAGGCTTCCGTTCTGAGGAGTAGATCTTGG 94655
Qy 221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla 240
Db 94656 GGAATTTCTTAACTGCTTTTGAAGGATGACATATGTATCTGAGACTTGTCTTATGGCC 94715
Qy 241 SerTyrTrrPrrIleAsnLysAsnValLeuThrAspProArgPheTyrAsnProGluGlu 260
Db 94716 AGTTTACGCGTGATCAATAAGACGATGACCGATCTCGCTTCTTAAATCCAGAA 94775
Qy 261 MetGluLysGlyLysLysGlyAlaLysProLysMetAsnMetLysAspSerPheLeuTyr 280
Db 94776 ATGCAAAAGGGAAGAAAGGAGTCAAACTTAATGAATGAATGAAGATACCTTCTTAT 94835
Qy 281 LeuAspArgSerProTrrIleLeuLeuThrLeuLeuValIleAlaTrrGlyIleCys 300
Db 94836 CTTGCAATGATCTCTTATATCTTATATGCTCTGTTGTTATGCTATGATTTGC 94895
Qy 301 IleAsnLeuIleGluValIleThrLysSerGlnLeuLysLeuGlnTrrProAsnMetAsn 320
Db 94896 ATTAACCTTAATCAAGACCTTGAAGAAAGTCAAGTCAATCAATCAATCAATCAAT 94955
Qy 321 AspTyrSerGluPheMetGlyAsnPheSerPheTrrThrGlyValaLysValLeuIle 340
```

| | | | | | |
|------------|--|---|-------|--------|---|
| Db | 94956 | GACATATGAGTTCATGGGAACTTCTCTTCTGAGTGGGTAGTATCCGACTTATTC | 95015 | gene | /isolate="GPIC" /db_xref="taxon:227941" complement (136. .1272) /locus_tag="CCA00275" complement (136. .1272) /locus_tag="CCA00275" /note="similar to GB:M19395, SP:P10655, GB:M28336, PID:143130, and PID:143134; identified by sequence similarity; putative" /codon_start=1 /product="polymorphic outer membrane protein E/F family protein, putative" /protein_id="AAP05026.1" /db_xref="GI:29834392" /translation="WSKLTFRYLTSTSLALGNVYADATPTTHTPTTKENDTKIDTQADIKETQTERETKTRDKGDFVANI FQOASVATTSGNNISRVCLDSLNNSFPDIEGGLALCLCYORXMEAGFPHDGAAYTELVSQSPSYKLGKFPASQRTAKANGHDEVASDYLSFSGSYMEVHLFEKGFILANGLYSRGLVHLNHTRKLIGACCSFSGTIGSALSFPFLKMIKNDRLTVMPEFRILAVLSHDPEEGARVPTSGSKPKDEPIDTSLPRGHSLAHTHGRFPSTWELVAYKSTQKQLPVSSKLVAADDTWTSTPNVSRRAVSVNLRNETQVLRKIMHVDYQCDLSSSTRSHYLAGRLSP" 1343. .1438 /locus_tag="CCA00276" 1343. .1438 /locus_tag="CCA00276" /note="identified by Glimmer2; putative" /codon_start=1 /translation="hypothetical protein" /product="hypothetical protein" /protein_id="AAP05027.1" /db_xref="GI:29834391" /translation="MGDLKSQSTNDPFWMEYIYILFIECVIL" complement (1670. .4615) /locus_tag="CCA00277" complement (1670. .4615) /locus_tag="CCA00277" /note="similar to GB:X65957, SP:P43148, and PID:396259; identified by sequence similarity; putative" /codon_start=1 /translation="polymorphic outer membrane protein H family protein" /product="polymorphic outer membrane protein H family protein" /protein_id="AAP05028.1" /db_xref="GI:29834392" /translation="WPSLRSSSFCLACICSSGVSAQTINQAKRPPLYDTSEAPLTNSCFEDYTGAKFPISFLGSSNFVPGNRSLIFQOASHAPVEGALISAVDSLKTDPCKLVNENISGRKGLISGNIAITSSKSLVITNNRAPYTPVTSVTPASTVNCFCGSAIHRETNLNTKIRLDNISFNSNSANFGAALIKTDATCTIENNAALISFCNFAACGGAIYDGTISIKNSGRLTLEGNANGLATTPPNVAVIAAGGALCAPTKSVPPANNSGICNTSYNLABODGAIYTTCDLATHATPYNNKNAATRYGALICAKALTINAHGRTIPVSNRAEKGAITYASTVGDITTAGFDSTLSSTSGDWVFVYNNMDSRGTRNAIDIEGGKLTSLNATLSIIIFHDPIDANGPVTVTLLDTTINSSCSCKVFTSETLSIEQENPNSTSLGKAVIEDGQLVVNNKAVNLGLAETGRLLDSGALIELYDNDLVNALISIPLAFTSNNGPSTLFTTGDLAVSKHYGQVMSWTIPIPLLAPTPNGISGCTNNRTLYVAMPDPAYRAPVAYAPRGRGLVNTLMTSLAQTQAEALERTISEQEGVILAKALIGSYVRHPMKTHGFPKQPGYQASLGHIPDDASVLAFOQLGVQVESKPYDAQSTEQISLVAFPGFPLATEVTVNISWEASGYAVNNMKTNVLAITQKTRQSGRWNNNTYVASVSVEHPLTMCULTRIAIDLEISGFIASFPWGMNAFSEOGALPRGSRGRGHNTITLPIGFTSEWGTFPKKPSMLTLKLAAYKPVYRVNPHNVVITLIANGERTPITGAQIPRHGFYLVQVHDSVELSRHATGFVDYVFDKSRVASHRITTGLOGRP" complement (4641. .7676) /locus_tag="CCA00278" complement (4641. .7676) /locus_tag="CCA00278" /note="identified by match to Pfam protein family HMM PF02415" /codon_start=1 /translation="polymorphic outer membrane protein G family protein/autotransporter" |
| Qy | 341 | MecleuphevalglylyasnvalilearglyspheglytrpLeuTh-GlyAlaleuVal | 360 | gene | |
| Db | 95016 | ATGCAATTTGTGTGGTGAACCTCATTCGTAAATTTGGATGGTTAACTGAGCCCTAGTC | 95075 | CDS | |
| Qy | 361 | ThrpVoleMecValleuLeuThrglylyleValPhePheAlaleuValilePheargan | 380 | gene | |
| Db | 95076 | ACTCTGTGATGTTCTCTCTCAACAGGATACGTTTCTTGCTCTGTATCTTAGAAGC | 95135 | CDS | |
| Qy | 381 | GlnAlaserGlyLeuValAlamerPheglyThrThrpProleuMecleuAlaValAlaVal | 400 | gene | |
| Db | 95136 | CAAGCTTGGGCTGTGCTGCTATGTTGGTCAACTCTCTCATCTAGCTGTGTTCTC | 95195 | CDS | |
| Qy | 401 | GlyAlaileglnAsnileuSerlySerThrlylyAlaleuPheAspSerThrlyls | 420 | gene | |
| Db | 95196 | GGAGCTTAAACAGAAATATCTTTCGAAATCCAAATATCGCTCTTCTGACTCAACTAA | 95255 | CDS | |
| Qy | 421 | GluMecAlaTyrlleProleuAspGlnGlnlyVallyVallyAlaAlaileasp | 440 | gene | |
| Db | 95256 | GAAATGGCTATATCCCTCTTGACCAAGACAAAGCAAGATGTAAGCTGCTATTTGAT | 95315 | CDS | |
| Qy | 441 | ValValAlaAlaArgPheglyLySerGlylyAlaleuileglnGlnlyLeuVal | 460 | gene | |
| Db | 95316 | GTAAGTGGCCGCCGCTTCGAAAAATCAGAGAGCTTTAAATCCAAAGATTGCTGCTT | 95375 | CDS | |
| Qy | 461 | IleCysglySerIleGlyAlaMecThrpTyrlleAlaValileleuPheilele | 480 | gene | |
| Db | 95376 | ATCTGTGAGATGATTTGAGAGCTATGACCCCTTATCTTGACGATTCCTTCTTCACTT | 95435 | CDS | |
| Qy | 481 | AlailetrpLeuValaserAlaThrlyleuAsnlyleuPheleuAlaGlnSerAlaleu | 500 | gene | |
| Db | 95436 | GCTATTTGCTTGTTCTTCGCAACTAGTTAAACAACACTATTTCTTAGCCAGTCTGCTCTT | 95495 | CDS | |
| Qy | 501 | LySGlnGlnlyValAlaGlnGlnlyAspSerAlaProAlaserSer | 515 | gene | |
| Db | 95496 | AAAGAACAGAAAGTGGCTCAAGAGATTCAAGTCTGCTTCTTCA | 95540 | CDS | |
| RESULT 6 | AE016995/c | 300512 bp | DNA | linear | BCT 14-APR-2003 |
| LOCUS | AE016995 | 300512 bp | DNA | linear | BCT 14-APR-2003 |
| DEFINITION | Chlamydomophila caviae GPIC section 2 of 4 of the complete genome. | | | | |
| ACCESSION | AE016995 | AE015925 | | | |
| VERSION | AE016995.1 | GI:29834389 | | | |
| KEYWORDS | Chlamydomophila caviae GPIC | | | | |
| SOURCE | Chlamydomophila caviae GPIC | | | | |
| ORGANISM | Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila. | | | | |
| REFERENCE | 1 (bases 1 to 300512) | | | | |
| AUTHORS | Read,T.D., Myers,G.S., Brunham,R.C., Nelson,W.C., Paulsen,I.T., Heidelberg,J., Holtzapple,E., Knouri,H., Federova,N.B., Carthy,H.A., Umeyan,L.A., Haft,D.H., Peterson,J., Beaman,M.J., White,O., Salzberg,S., Salzberg,S.L., Hsia,R.C., McClarty,G., Rank,R.G., Bavoiil,P.M. and Fraser,C.M. | | | | |
| TITLE | Genome sequence of Chlamydomophila caviae (Chlamydia psittaci GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiaceae | | | | |
| JOURNAL | Nucleic Acids Res. | 31 (8) | | | 2134-2147 (2003) |
| MEDLINE | 12682364 | | | | |
| PUBMED | 12682364 | | | | |
| FEATURES | 2 (bases 1 to 300512) | | | | |
| REFERENCE | Read,T., Myers,G., Brunham,R., Nelson,W., Paulsen,I., Heidelberg,J., Holtzapple,E., Knouri,H., Federova,N., Carthy,H., Umeyan,L., Haft,D., Peterson,J., Beaman,M., White,O., Salzberg,S., Hsia,R.-C., McClarty,G., Rank,R., Bavoiil,P. and Fraser,C. | | | | |
| JOURNAL | Submitted (29-OCT-2002) The Institute for Genomic Research, 9712 | | | | |
| TITLE | Medical Center Dr, Rockville, MD 20850, USA | | | | |
| LOCATION | Location/Qualifiers | | | | |
| FEATURES | 1. .300512 | | | | |
| SOURCE | /organism="Chlamydomophila caviae GPIC" | | | | |
| | /mol_type="genomic DNA" | | | | |

```
/protein_id="AAP05029.1"
/db_xref="GI:29834393"
/translation="MKNSLRKFLITLTLTPYSPQASLEFVNVNGTYDGLRETPPY
TITNSPEPTALISGNLNLNLNLSNMAVTPSSCFRNSAGSTTVGRNHNLTFTLNRIS
ANGALSTIPPTPESEFPYTIKGVNTLSFNSCLMAMATTPATTPVANGAFAFS
KAPFLENIQVLFKNNRAADSGGLLWETLSTINIKSMQPLSNVANGAIAKASG
LDVQPCSEILFRNSAERLQGAIOAVDPATNOVTAVRFSEGVTPDQDANAGGA
IYSKGVDFSNNAQOLLIONNSAPENVANTNVLGGGAIIFCVQGTPOPPPPPTPT
PFSGGLTTTNOKKOLLFANNPATAGAYGKESITSGCKPTNNIATKCGAIYIBE
NGELTSLADYDGMFYENLKQDQATVTNNATTLAKGATIKLLASGDKKCPPIYV
TLPEAPNDKTLTINODKTSSTPTNTIGLPSGAVDSQASATNASTTYQRY
ILGGKVLADKASLSVASFQETDSILMDNGTLLAITESHQTPAAGGGGGGTP
TOEANTGVISLTNLHVNISLTQGEBAKLETNKTDTITLTGHSILDDVSGAYEN
HDLNKTQVITNLISLTPAGSKTTINGLDILRGDAPVYQSGSMDLAKEMENDAN
KOKILAKMTKTGTSTNPEROSALVNSLKGAFIDLRSMNALATASCDGPEYGVAM
AGISNFRHNRNSVSHGFRISGCVIGANSQTYDVGFVAFSOLFAKSKDYVSSA
KSQALASAYLSVRQULNTIIFSSPAARINSHNEDMKTYTFIPEKDQWDMCWL
GEIGSLPIVQITKHLNQTIPKMNQOLGAHSGPEKLEARSFCSRLINLAVP
VGFKIDRRSHDPFYSLSIYIPDMVRNPGCNTLLANGVRWKTPTATNLNRGLLM
QGSTHTAVLSNIEIFSHSGCELRSSRNRYNINVGSKIRF"
complement(7925..10531)
/locus_tag="CCA00279"
/locus="similar to GB:M25260, SP:P26778, and PID:304172;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic outer membrane protein G family
protein"
/protein_id="AAP05030.1"
/db_xref="GI:29834394"
/translation="MTQRPMPWIFCCCLSCLSQSPAAVLTUKLTATNDPDGINNA
PRTTSDSECTNYLINDVITQNVATNPNSCFKATKCNLFENGNNRSLTNNTT
TAVGKALINSSEFELTSMSCSKQLQLEPVSSRKQSLVSKGLTMRNELLFDGC
VYSKQGAICMPSPGARNSLTFPSNRKATFENNNSVNGGAIYQHLIYIAGPT
LFTNNMSGDVNSRPGGALAIADQELSLPADQDITFSGNLASRGHFPKNAIHE
NDAYISYITAKEGSVTFYDAITSNASTVTLINOLTLDGTHGSAVLAKGCVASG
SLISTMSRPAIDQITLLAGSINSGVILTNSTIQHPNTPTPIDQSTMLRQENH
ILNNTIPVKQTLVLRPKESRKRNLSRSRPGQITLSQDKITLSLSQINIDDED
FYENLELANPVKQTLIHEARELEDIRKIDLHPVNPVYQSGMDLQWMDTFSQOI
STRILANAEKTAQQLWTPKHYRPLIDQGSILVNSIMNFAVIRGIHNMIDVDSG
SLYNGLMAAEVSNVFKHDKHNRGPFHTGGVALISQOTPEKDIFFSLQFPMFG
SKDASLAKNHEIILAGSYVAQHSIIPVQILWMSIGNMWRPEPFSKISDPLINQ
ATYHSKKNKLTISKSSRITGTGMYNTHGTGELSSISFPDEEHNFMHILPMMNQ
GVAVYQRFKESGKKHETSSKLGNSLAPGIRLEGHTSCWPIFVATSVAFIADLYR
QNPCTITATYEPALIWETSGTNERQALNQLNSHCALLENITLFSKCAEIRKSSQ
YSGDISOFLF"
complement(10816..13347)
/locus_tag="CCA00280"
/locus="similar to GB:S70840, and PID:546752; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic outer membrane protein G family
protein/autotransporter, putative"
/protein_id="AAP05031.1"
/db_xref="GI:29834395"
/translation="MKNSLVGLIFSSFTASIAFHSVNAADNLNLSSESFDGSGAGF
TSKQSDAGCTTYLTGDDVTITVHTKTSNPATSCPKNSGTLFTGKANHSLFEDIS
TRQGAISANTDAKTLTKSGPNTLAFVAPATTTGNAAYVATGTTTKENKALFGKN
HSTAVGAIHCAKGAITEAVLTLOONASMLFRNNSATGAIHAEKRLIAGQGTTLF
ENNAHQKGAISIASGSEISLSADDSIIFKQNTIDAGGVNGIHVNGKFAVL
EAKSOSILFYDPVAVAGTADSNLEINKADGTSYGSIIIFSGRYSIPHTGMRKAVT
FTQPLTSSGLVLEKGAHLKAKSLVOTAGSKVLLDQTSIETKENIDIKELMLRDL
FNPTATCISGNAHTTIKGLPGLVDGTFYDNALAYSIDOEFLQADKQITKI
SLVDIPOAVRNKLDHSGYQCKMSIDKTVGSRNACVTTTGTTATVTHMRPTIYR
GGSGEITPLVAVNTKGNFSDIRNLERVELANSICSBGFMAAGIKNYIYSNPAV
NYVQHNAGVAGIMNKRHTISENVSAFSQLFKDHDHANGVHDQTLTSGSFYAHV
GSLPMLRFLCGSGSCPPELOQASPSIPVIAVALQSYSHNNHLTIHEDTKTKGMS
NYSIAELGTFVYTLKCPISILKNVSPFVQLGVYSEORFSEBGRRCLEFSTYLA
```

```
gene
cDS
NLAFLPIKIHGVCRELFAVDLSAMTVHDFRIDPMTMLFLIGLA PMWTHANNILA
TKAIVQSGRFAFRSNI EYFAEBCNCLRSSHSYNDFGAKIHF"
complement(13501..16032)
/locus_tag="CCA00281"
/locus="similar to GB:S70840, and PID:546752; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic outer membrane protein G family
protein/autotransporter, putative"
/protein_id="AAP05032.1"
/db_xref="GI:29834396"
/translation="MKNSITGVLFSSFALSIAETELLADADTVNLAFRNSSTEN
VKQNNVGEITTYITLGAVSFANINKFQDQDLSCRANSAGDITFGSRRLTFNNITSS
AKGAISTTDADKTLTSLGCLSLFVMSPEIDQNGAIYNSNMLIIONSVSGYK
SAGKSVILCEKTSAGATPLTIOUNNGERVLANLSASGGAIVAEKMLSAGNT
VFOANVOBKGAIAIANGESISLADGDLTFERNITIEENETIRNAHLEDQAKFL
OLRAKRSIYFHPDITGVNADRLTLNANGATPYEGTIVFAFSVYNSPLSKLY
SFSQDUTLNASLILKDVYLIKAKSPONOKSLFPMRGTRLQTTNNISIKHLIDS
KIATVAEIAATVDAALIEICGPMWVHDEIFYNQALDLSFECULVAPHDNI
```

| Alignment Scores: | | | | | |
|---|---|---------------|--------|--|--|
| Pred. No.: | 6,48e-187 | Length: | 300512 | | |
| Score: | 2242.50 | Matches: | 424 | | |
| Percent Similarity: | 92.02% | Conservative: | 49 | | |
| Best Local Similarity: | 82.49% | Mismatch: | 40 | | |
| Query Match: | 85.27% | Indels: | 1 | | |
| DB: | 1 | Gaps: | 1 | | |
| US-09-869-433-2 (1-515) x AE016995 (1-300512) | | | | | |
| QY 1 | MechThystrhguglulyspPhehglylsleuAysSerPheleutPpcoilhis 20 | | | | |
| Db 188695 | ATGACACAAACAGCGGAAAAACCTTTGGAAATTTGGCGTCTTTCTTGCCAAATCAT 188636 | | | | |
| QY 21 | ThrihgiuleuylsylvallleuPmetPheleuuetPhepheySilleThPhasn 40 | | | | |
| Db 188635 | ATGCATAGCTGAAAGAGGCTGCGCAATGTTCTTAATGTTCTTGATATTCGCTTTAAC 188576 | | | | |
| QY 41 | TyThrValleuAysaPThrlyAsPThrleuileValgIyAaPProglYserGlyVala 60 | | | | |
| Db 188575 | TACACGATATTCCTGATACAAAGACACTTTATCTGAACGCTCCCGGCTCCGGTCA 188516 | | | | |
| QY 61 | GlulallepPhepHeileuysPheTrpLeuValaProCyalaIleilePheMetleu 80 | | | | |
| Db 188515 | GAGGCTADACCTTCATCAATTAATGCTGCTGCTGCTGCTGCTGCTGCTCATGCTT 188456 | | | | |
| QY 81 | IleTyThralysleuSerAenilleuSerilyseGlnAlaleuPProglYleuGlyleu 100 | | | | |
| Db 188455 | ATCTATGCCAAGCTTAACCAATATTTTAAAGAACGACCCCTCTTACACAGTCATCAC 188396 | | | | |
| QY 101 | ProPheleuilePhepHealaleuPheProThrValIleTyThrProleuAysValleu 120 | | | | |
| Db 188395 | CCATTTCTTTATTTCTTCCTGCTTATTCCTTACAGTGAATTTTACATTCCTGCAATCTTA 188336 | | | | |
| QY 121 | HiePThrgiuphealAasPArgleuGlnAlaleuPProPoglyleuGlyleu 140 | | | | |
| Db 188335 | CACCAACGATTTTGGCGAATTAATTAACACCATCTTCCTGAGGGGTTGATGGGTTGT 188276 | | | | |
| QY 141 | ValAlalleuAysaenITPThPhaAlaAlaheTyValleuAlaGlyleuTrgly 160 | | | | |
| Db 188275 | GTACCTATGCTTAAACAACTGACCTTTCGACGCTTTTACGCTTTCGACCTATGGGGA 188216 | | | | |
| QY 161 | SerValmetleuSerleuMetPheTrpGlyPheAlaasnGluIleTrlySileGly 180 | | | | |
| Db 188215 | AGCGTATGCTCTCTTAATGCTTCTGGGCTTTTCTGTAATGAATTAACAAATTAAGGGA 188156 | | | | |
| QY 181 | AlAlaysArgPheTyAlaaleuPheGlyIleGlyValaasnIleSerleuLeuAlaSerGly 200 | | | | |
| Db 188155 | GCTAAACGTTTCTACGCTTATTCGAGTGTGCTCAACATCGCTTACTAGCTTCCGGA 188096 | | | | |

```

Qy 201 ArgAlaIleValITrpAlaSerIysLeuArgAlaSerValSerGluGlyValAspProTrp 220
Db 188095 CCGTCTATTACTGGGCTTCTTAAGCTACCTGCTAGCGCTACAGGAAATACATCTTGG 188036
Qy 221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla 240
Db 188035 GGACCTTCCCTTACCTCTTAATGAGTATGTTATCTTCCGGCGGCTTAATTTCTTA 187976
Qy 241 SerITyTTrpTrpIleAsnIysAsnValLeuThrAspProArgPheTyTrpAsnProGluGlu 260
Db 187975 TGCTACTGGTGGATGAAATAGCGCTTCTTACAGATCTCAAGATCTTAAAGAA 187916
Qy 261 MetGlnIysGlyIleLeuGlyValAlaIysProIysMetAsnMetIysAspSerPheLeuTy 280
Db 187915 TTGAACAAATCCAGAAA---TCCAAGCCAAATAATGAGATAGAGAAAAGCTTCTTAT 187859
Qy 281 LeuAspArgSerProTyTrpIleLeuLeuLeuThrLeuLeuValIleAlaTyGlyIleCys 300
Db 187858 TTAGCAAGATCTCTTACATGTTATTTACTAGCCCTTCTTGTATTTGCTACGGCTTGT 187799
Qy 301 IleAsnLeuIleGluValITrpTrpIysSerGlnLeuIysLeuGlnTyTrpProAsnMetAsn 320
Db 187798 ATTAACCTCGTAGAAGTTAACCTGGAAGCCAAATGAAATGCATACCAATCTTAC 187739
Qy 321 AspTyTrpSerGluPheMetGlyAsnPheSerPheTrpThrGlyValValSerValLeuIle 340
Db 187738 GAATATAGCCAGTTATATGGTAACTTCTCATTTCTGACCGGGGCTTATCCTTATCGG 187679
Qy 341 MetLeuPheValGlyIleAsnValIleArgIysPheGlyTyTrpLeuThrGlyValLeuVal 360
Db 187678 ATGTTATTTGTCGGTGTACCTTAATCCGTTAATCGTTGTTAAGAGAGCTCTCGTA 187619
Qy 361 ThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn 380
Db 187618 ACTCCAGTATGATGATGTTAAACAGCGCTTCTTCTTCTGCTGATATCTTCAGAGAT 187559
Qy 381 GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValVal 400
Db 187558 CAGGCTTCTGAGATGCTAGCTACGTTAGGAACACCACTTAATGCTTCTTATGGT 187499
Qy 401 GlyAlaIleGlnAsnIleLeuSerIysSerThrIysTyTrpAlaLeuPheAspSerThrIys 420
Db 187498 GGGGCTATCCAAAATCTCTATCTAATCCAGAAATACGCCCTCTCGATCAACAAA 187439
Qy 421 GluMetAlaTyTrpIleProLeuAspGlnGlnIysValIysValGlyValAlaIleAsp 440
Db 187438 GAAAGGCTTACATTCATGATGATCAAGAACAAAAGTTAAAGAAAGCTGCTATGAC 187379
Qy 441 ValValAlaAlaArgPheGlyIysSerGlyIysAlaLeuIleGlnGlnIysLeuVal 460
Db 187378 GTAGTTGCTGCTGCTTTCGAAAAATCCGGCGCTTCTTATATCCAAAGTCTTCTAGTG 187319
Qy 461 IleCysGlySerIleGlyAlaMetThrProTyTrpLeuAlaValIleLeuLeuPheIleIle 480
Db 187318 GTTGTGGAAGATTTGAGCGCATGACTCTTACCTAGCAATGACTCTTCTTGTATTT 187259
Qy 481 AlaIleTrpLeuValSerAlaThrIysLeuAsnIysLeuPheLeuAlaGlnSerAlaLeu 500
Db 187258 GCTGATGCTTATGCTCAGCTACAAAACCTTAATTAAGCTTCTTAACTCAATGCGCATC 187199
Qy 501 LysGlnGlnIysValAlaGlnIysAspSerAlaProAlaSer 514
Db 187198 AAAGAACAAAGAACTTGCAGAGAGAAAGCGCTGCTCTTCT 187157

```

```

RESULT 7
AB001281/c 14168 bp DNA linear BCT 30-OCT-2000
LOCUS Chlamydia trachomatis section 8 of 87 of the complete genome.
DEFINITION
ACCESSION AB001281 AB001273
VERSION AB001281.1 GI:3328454
KEYWORDS
SOURCE Chlamydia trachomatis
ORGANISM Chlamydia trachomatis

```

```

REFERENCE
AUTHORS
Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
1 (bases 1 to 14168)
Stephens, R.S., Kaiman, S., Lammel, C.J., Fan, J., Marathe, R.,
Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,
Koonin, E.V., and Davis, R.W.
Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis
Science 282 (5389), 754-759 (1998)
JOURNAL
MEDLINE
99000809
PUBMED
9784136
AUTHORS
2 (bases 1 to 14168)
Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,
Olinger, L., Grimwood, J., Davis, R.W., and Stephens, R.S.
Comparative genomes of Chlamydia pneumoniae and C. trachomatis
Nat. Genet. 21 (4), 385-389 (1999)
JOURNAL
MEDLINE
99206606
PUBMED
10192388
AUTHORS
3 (bases 1 to 14168)
Stephens, R.S., Kaiman, S., Lammel, C.J., Fan, J., Marathe, R.,
Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,
Koonin, E.V., and Davis, R.W.
Direct Submission
Submitted (20-May-1998) Program in Infectious Diseases, University
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
FEATURES
source
1..14168
location/Qualifiers
/organism="Chlamydia trachomatis"
/mol_type="genomic DNA"
/strain="D/UW-3/CX"
/db_xref="taxon:813"
86..847
/gene="flaA"
/note="synonym: CT061"
86..847
/gene="flaA"
/codon_start=1
/transl_table=1
/product="Sigma-28/WhiG Family"
/protein_id="AAC67652.1"
/db_xref="GI:3328455"
401..2244
/transl_table=1
/product="tyrosyl tRNA Synthetase"
/protein_id="AAC67653.1"
/db_xref="GI:3328456"
/transl_table="MOOLINDLKKRGILDNSSAGLESLTYVPSAYLGFDPAPSLHIG
HTMIGICIRLRLAAGITPVALVGATGATGIDPSPGKSVSRSLDQALDNSSKAAAL
ASYIPGIRYVNMADMLGSLMVDPLRVGCHFGSLADLVKORVYSESGISYTF
SYLLQSYDTPAHLPKEHNVVLCQGSQDMGNTSGIDYIRRGIGQAYGLTPLYLTS
KGGKIGTRESGITWLDPALTPPELFOYFRLPDEISKWRITLTDNEIEFLDER
LTSPPQVAKYIAEVIKDVHSGELAQAAATESFPASGKSTTEBELVALVSGVG
VKVARADLIGKRMIDIVELGFCSSRGQARRLIQORGLYINQEPLADEQSLDGTQC
FDRVYLISGKKRKKQVIDLN"
2254..3696
/gene="gnd"
/note="synonym: CT063"
2254..3696
/gene="gnd"
/codon_start=1
/transl_table=1
/product="6-Phosphogluconate Dehydrogenase"
/protein_id="AAC67654.1"
/db_xref="GI:3328457"

```



```

/locus_tag="TC00336"
/note="Similar to GB:M63957, SP:P30720, PID:151242, and
PID:999193; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAP9199.1"
/db_xref="GI:7190378"
/translation="MNRQTQTTQEBLEQLTASVGGSKKNDLCRIPEPTGGIVH
HFLTKKSHSPQLMLKLTPTVIDSTPTSPINPKRAARGSKKKDVPNFTKIDIE
VLEIARQGVDDLDLAPSPKPSKRELIIRSNINIVSTELMAVYEAALSOPE"
complement(10169..10270)
/locus_tag="TC00337"
complement(10169..10270)
/locus_tag="TC00337"
/note="Identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAP9200.1"
/db_xref="GI:7190379"
/translation="MQNINLNNLKHPIFCPPNOKISKVRLKKQKPL"
10658..11638
/locus_tag="TC00338"
10658..11638
/locus_tag="TC00338"
/note="Similar to GB:X63368, GB:X63369, SP:P25686,
PID:32469, and PID:32470; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1

```

Alignment Scores:

| | | | |
|------------------------|----------|---------------|-------|
| Pred. No.: | 1.8e-179 | Length: | 23773 |
| Score: | 2143.00 | Matches: | 408 |
| Percent Similarity: | 89.04% | Conservative: | 55 |
| Best Local Similarity: | 78.46% | Mismatches: | 49 |
| Query Match: | 81.48% | Indels: | 8 |
| DB: | 1 | Gaps: | 3 |

US-09-869-433-2 (1-515) X AE002301 (1-23773)

| | | | |
|----|------|--|------|
| QY | 1 | MetThrIysThrGluGluValProPheIleLysLeuAspSerPheLeuTrpProIleHis | 20 |
| Db | 9206 | ATGATCCAAACCGCGAAAAACCTTTTGAAAATTGGCCCTCTTCCCTGGCCGATAC | 9147 |
| QY | 21 | ThrIleGluLeuLysValLeuProMetPheLeuMetPhePheCysIleThrPheAsn | 40 |
| Db | 9146 | ATGATGATAGCTTAAGAAAGTTCTGGCAATGTTCTTAATGTTCTTGTATTTTCATTCAAT | 9087 |
| QY | 41 | TyrThrValLeuAspArgSerPThrLysAspThrLeuIleValGlyAlaProGlySerGlyAla | 60 |
| Db | 9086 | TACACGATTTTGGAGGATACAAAAGACACTTTATCGTACAGACACGAGCTCTGGACGA | 9027 |
| QY | 61 | GluAlaIleProPheIleLysPheProLeuValValProCysAlaIleIlePheMetLeu | 80 |
| Db | 9026 | GAGCGCATTCCTTTATTAAGTTGGCTGGTGTCTTCTGCTGGTGGTCAATGCTT | 8967 |
| QY | 81 | IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr | 100 |
| Db | 8966 | ATCTACGGCAAAGTTAAGCAATATTTTGGACAAACGCGCTTTTCTATGACAGTGCTCT | 8907 |
| QY | 101 | ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuAspValLeu | 120 |
| Db | 8906 | CCATTCTGTTGGTTTTTGGACGTGTTCCCTTGGATTTACCCATATGTCATATACCT | 8847 |
| QY | 121 | HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu | 140 |
| Db | 8846 | CACCCAAAGACCTTCGTGATACACTAAGCGCATTCCTCTCGGGTTTCTAGCGTTC | 8787 |
| QY | 141 | ValAlaIleLeuAspGluTrpThrPheAlaAlaPheTyrValLeuAlaGluLeuTrpGly | 160 |
| Db | 8786 | ATTTCGATGATACGCACTTCGACTTTCACAGGTTTTATGTCTTTCCGAACCTTTGGGGA | 8727 |

| | | | |
|----|------|---|-----|
| QY | 161 | SerValMetLeuSerLeuMetPheThrGlyPheAlaSerGluIleThrLysIleHisGlu | 180 |
| Db | 8726 | AGTGTAAATGCTTGGCTTAATGTTCTGGGGATTGGCAATGCAATCACTAAATTTGCGGA | 866 |
| QY | 161 | AlaLysArgPheTyPAlaLeuPheGlyIleGlyAlaSerIleSerLeuLeuAlaSerGly | 200 |
| Db | 8666 | GCTTAAACGTTTCTACGCTCTTTTGGAGTTGAGCTTAACGTAAGTACTCTTAATTTCCGCT | 860 |
| QY | 201 | ArgAlaIleValTPAlaSerLysLeuArgAlaSerValSerGluIValaAspProTrp | 220 |
| Db | 8606 | CCACCTAATGTTGGTCTCTTAATTAAGCCCAAGCTTAAGTGAAGCGGTAAACCATG | 854 |
| QY | 221 | GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla | 240 |
| Db | 8546 | GGAGTTACTCTTACTCTTAATGAGCTAATGTTCCGTTCTTTGGCATTAATCCCTGCC | 848 |
| QY | 241 | SerTyTPTrpTPIleAsnLysAsnValLeuThrAspProArgPheTyAspProGluIu | 260 |
| Db | 8486 | TGCTAATCGTGGTAAGAAATCGTTAGCTTCTAATCTTAAGTTCTTAATTAACCAACAGAG | 842 |
| QY | 261 | MetGlnTyPegGlyLysLysGlyAlaLysProLysMetLysAspSerPheLeuTyR | 280 |
| Db | 8426 | CTA---AAAGTAAAAA---TCTTAACCTTAAGTACAGACATGGCGAGAGCTTCTTAT | 837 |
| QY | 281 | LeuAspArgSerProTyRLeuLeuLeuLeuThrLeuLeuValIleAlaTyRglyLecys | 300 |
| Db | 8372 | CTGTTAAGATCCCTTACATCTCTTCTTACGCTTGAATTAATTTGGCTAACGAGTTTGC | 831 |
| QY | 301 | IleAsnLeuIleGluValThrTriPysSerGlnLeuLysIleuGlnTyRProAsnMetAsn | 320 |
| Db | 8312 | ATCAACCTGTTGGAACTACTTGGAAAGCCAGCTCAAAATGCAATTCCTTAATCCAAAT | 825 |
| QY | 321 | AspTyRSerGluPheMetGlyAsnPheSerPheTrpThrGlyValaIleSerValLeuIle | 340 |
| Db | 8252 | GAATACAGCGCATTCATGGGAACATCTCTTCTTGGACAGGGGTTGTATCCGTAATTCGT | 819 |
| QY | 341 | MetLeuPheValGlyGlyAsnValIleArgLysPheGlyTriPLeuThrGlyAlaLeuVal | 360 |
| Db | 8192 | ATGCTGTTCATTTGGTGGTATGTTATTCGCTGTTTGGCTGGTTAAACAGGCCCTTAGTT | 813 |
| QY | 361 | ThrProValMetValLeuLeuThrGlyLysIleValPhePheAlaLeuValIlePheArgAsn | 380 |
| Db | 8132 | ACACCTGTCAGAGCTCCCTGTAACAGAGCGCAATCTTGGCTCTCGTAATCTTCAGAGAT | 807 |
| QY | 381 | GlnIleAserGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValaVala | 400 |
| Db | 8072 | CATCAACAGGATTTGGTTCAGCTTTGGGAACAATCCTTTAATGCTACGGGTTGT | 801 |
| QY | 401 | GlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyPAlaLeuPheAspSerThrLys | 420 |
| Db | 8012 | GGAGCCCGTCCAAATATTCTTCCAAATCGACTAAATATGCTCTCTTTGATGCAACTAA | 795 |
| QY | 421 | GluMetAlaTyRleProLeuAspGlnGluGlnLysValLysGlyLysAlaAlaIleAsp | 440 |
| Db | 7952 | GAGATGGCTTACATCCCATTCGATCAAGACAAAGAAAGTTAAAGAAAGCCGCTATTGAC | 789 |
| QY | 441 | ValValaAlaAlaArgPheGlyLysSerGlyLysAlaLeuIleGlnGlnGlyLeuLeuVal | 460 |
| Db | 7892 | GTTTGTGTGCTCAATTTGTTAAATCTGGGGGATCTTTGATTCCAACAAAGAACTCCTAGT | 783 |
| QY | 461 | IleCysGlySerIleGlyAlaMetThrProTyRLeuAlaValIleLeuLeuPheIleIle | 480 |
| Db | 7832 | GTTTGGCGAAGCATTCAGTGCATATCTCATTTCTTACGGGTATGCTCTTTGGCAATTAT | 777 |
| QY | 481 | AlaIleTriPLeuValSerAlaThrLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeu | 500 |
| Db | 7772 | ATGTTTGGTTAAACATCTGCAACCAACTTAATTAATTTGTTCTTAGCTGCTTCACTGCT | 771 |
| QY | 501 | LysGluGlnGluVal-----AlaGlnGluAspSerAlaProAlaSer | 514 |
| Db | 7712 | AAAAAACAAGCTGGCGACAGGCAAGCAGACAGCTGGAAGGAAGCTTTCTCTTCAGCG | 765 |

[illegible]

KEYWORDS
SOURCE Oryza sativa
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
1 Glazebrook, J., Briggs, S., Cooper, B., Goff, S.A., Moughamer, T., Katagiri, F., Kreps, J., Provart, N., Rhee, D. and Zhu, F.
Plant disease resistance genes
Patent: WO 0300906-A 743 03-JAN-2003;
JOURNAL Syngenta Participations AG (CH)
FEATURES
source location/Qualifiers
1. 1851
/organism="Oryza sativa"
/mol_type="genomic DNA"
/db_xref="taxon:4530"

BASE COUNT 401 a 453 c 509 g 487 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 4.81e-113 Length: 1851
Score: 1379.50 Matches: 275
Percent Similarity: 71.71% Conservative: 90
Best Local Similarity: 54.03% Mismatches: 135
Query Match: 52.45% Indels: 9
DB: Gaps: 6

US-09-869-433-2 (1-515) x AX660386 (1-1851)

Qy 7 LysProhegilyLysLeuArgSerPheLeuThrProIleHisThrHisGluLeuLys 26
Db 235 CAGCGGAGGGGAGG-----AAGTTCTCT--GGCGTGAAGGTGAAGACCTTAAGAAG 285

Qy 27 ValLeuProMetPheLeuMetPhePheCysIleThrPheAsnYrThrValLeuArgAsp 46
Db 286 ATCGTCGCGCTCGGGTGGATGTTCTTCTGCACTTGTTCACATCAAGATCTCGGGAC 345

Qy 47 ThrLysAspThrLeuIleValGlyAlaProGlySerGlyAlaGluAlaIleProPheIle 66
Db 346 ACCAAGACGTCGCTGGTACCGCCAGGGGAGCGCGGAGATCATCTCCCTTCTC 405

Qy 67 LysPheThrLeuValValProCysAlaIleIlePheMetLeuIleYrAlaLysLeuSer 86
Db 406 AAGACGTTGGGCAACCTGCCATGCGCTTCACTTCCCTTCAACCAAGCTTCC 465

Qy 87 AsnIleLeuSerIlyGlyAlaLeuPheYrAlaValGlyThrProPheLeuIlePhePhe 106
Db 466 AATGTGCTCTCAGGAGGCGCTCTTCAACCGTCACTTCCCTTCAATCGCTTCTC 525

Qy 107 AlaLeuPheProThrValIleYrProLeuArgAspValLeuHisProThrGluPheAla 126
Db 526 GGCCTCTTGGCTTCGCTCTACCTCTCAGAAATGATGATCAACCCACGCGCTCGCC 585

Qy 127 AspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeuValAlaIleLeuArgAsn 146
Db 586 GACAACTCTCTGGCGGCGCTCGCCGAGCTTCTTGGCCCGCTTCAAGATTT 645

Qy 147 ThrThrPheAlaAlaPheYrValLeuAlaGluLeuThrGlySerValMetLeuSerLeu 166
Db 646 TGGAGCTTTCCTTCTTATGTCATGCGGACGATGAGGCGACGCTGATCTCGGTC 705

Qy 167 MetPheThrGlyPheAlaAsnGluIleThrLysIleHisGluAlaLysArgPheYrAla 186
Db 706 CTCCTTCGGGCTTCGCTTCAACAGATTACTACGTTGAGGAAGAAAAGATTCTACCT 765

Qy 187 LeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGlyArgAlaIleValThrAla 206
Db 766 CTATTTGGCTTGGGCGCTTAATATGCGCTCACTTCTTCTGGGCGTACTGTAAGATTTC 825

Qy 207 SerLysLeuArgAlaSerValSerGluGlyValAspProThrGlyIleSerLeuArgLeu 226
Db 826 TCAAATCTGGCCAGACATTGGTCCAGGTATTGACGATGGAGGATGCTTTTAAAGA 885

Qy 227 LeuMetAlaMetThrIleValSerGlyLeuValLeuMetAlaSerYrTrpTrpIleAsn 246
Db 886 ATATGATGCTTAGTGGGCTTCTTGGACTGTGATCACTTCCATCTATGGGGGTAAC 945

Qy 247 LysAsnValLeuThrAspProArgPheYrAspProGluGluMetGlnLysGlyLys 266
Db 946 AAGTTGTTTGAATATATCTTCTCTT-----CCGAAAGCTGACCGTAAAGAAAAG 999

Qy 267 GlyAlaLysProLysMetAsnMetLysAspSerPheLeuYrLeuAspArgSerProYr 286
Db 1000 GAT---AAACCTAAGCTTGAATGAAGAAAGAGCTTAAAGTCTCTCTCTCAAGATAT 1056

Qy 287 IleLeuLeuLeuThrLeuLeuValIleAlaYrGlyYrIleCysIleAsnLeuIleGluVal 306
Db 1057 GTAAGGACCTTGCTACTTATGTTGGTGAATGACATCAAGTATCAACCTTGGAGCTC 1116

Qy 307 ThrTrpLysSerGlnLeuLysLeuGlnYrProAsnMetAsnAspYrSerGluPheMet 326
Db 1117 ACATGAAATCGAAGCTCAAGGACAGCACTTCCCTAGTCCAAATGATATCTTTCATG 1176

Qy 327 GlyAsnPheSerPheThrGlyValValSerValLeuIleMetLeuPheValGlyGly 346
Db 1177 GGTGATTTCTCAACCGCTACTGGCAATGCTTCACTTCAATGATGCTG--TTAGCCCGG 1233

Qy 347 AsnValIleArgLysPheGlyYrThrPheThrGlyAlaLeuValThrProValMetValLeu 366
Db 1234 ATCATACTCGAAATATTCGCTGGAGAGTGGCTGCATGATCAACCCACAGTTTGTG 1293

Qy 367 LeuThrGlyIleValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuVal 386
Db 1294 CTGATGCGCTCGGTTCTTCTTCCCTGATTTCTATTGGGACGCTCATCTCCATGCTT 1353

Qy 387 AlaMetPheGlyYrThrProLeuMetLeuAlaValAlaGlyAlaIleGlnAsnIle 406
Db 1354 GCCAATATGGGCAATGACCCCTCTTCTGACAGATTATGATGGTGACATTCAAGACATA 1413

Qy 407 LeuSerLysSerThrYrYrAlaLeuPheAspSerThrYrGluMetAlaYrIlePro 426
Db 1414 TTGAGCAAGAGTCCAAAGATATGATTTATTCATCCATGCAAAAGAGGCTTCAATTCCT 1473

Qy 427 LeuAspGlnGluGlnLysValLysGlyLysAlaAlaIleAspValAlaAlaArgPhe 446
Db 1474 TTGATGAAATATATGAAAGTGAAGGAGGAGGAGCCATTATGTTGCTTCCAAACCTTTC 1533

Qy 447 GlyLysSerGlyGlyAlaLeuIleGlnGlnGlyLeuLeuValIleCysGlySerIleGly 466
Db 1534 GGGAACTGCTGGAGCTTATTCACAGTTCATGATCTTCAATTGCTCTTTCAGCT 1593

Qy 467 AlaMetThrProYrLeuAlaValIleLeuLeuPheIleIleAlaIleThrPheValSer 486
Db 1594 AACTCCACCCCTTACCTTGGTGCATACCTGATGATTTCTTTCAGCATGGCTTACGGCGC 1653

Qy 487 AlaThrLysLeuAsnLysLeuPhe-----LeuAlaGlnSerAlaLeuLysGluGlnGlu 504
Db 1654 GCAAGTCACTTACACAGCAATCTCTAGCTTGGCTTGAAGAAATCTCAAAAGGATATG 1713

Qy 505 ValAlaGlnGluAspSerAlaProAla 513
Db 1714 TCTGCAAAAGAGAAAGTATGATCTTCC 1740

RESULT 12
AX654737 1872 bp DNA linear PAT 22-MAR-2003
LOCUS AX654737
DEFINITION Sequence 4607 from Patent WO0300898.
ACCESSION AX654737
VERSION AX654737.1 GI:29157551
KEYWORDS
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1
 AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
 Karaganti, F., Qian, S., Tao, Y., Whitman, S., Xie, Z., Zhu, T. and Zou, G.
 TITLE Plant genes involved in defense against pathogens
 JOURNAL Patent: WO 0300898-A 4607 03-JAN-2003;
 Syngenta Participations AG (CH)
 FEATURES Location/Qualifiers
 SOURCE 1..1872
 /organism="Oryza sativa"
 /mol_type="genomic DNA"
 /db_xref="taxon:4530"
 BASE COUNT 384 a 545 c 497 g 442 t 4 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,14e-109 Length: 1872
 Score: 1338.50 Matches: 266
 Percent Similarity: 70.38% Conservative: 88
 Best Local Similarity: 52.88% Mismatches: 141
 Query Match: 50.89% Indels: 9
 Gaps: 3
 DB: 6
 US-09-869-433-2 (1-515) x AK654737 (1-1872)

Db 958 -----:TAGCTGGCCAGCTGCTGAGCGGAAAGAAAGATAGCCAAACTG 1005
 Qy 273 AsmMetLysaSerPheLeuTyrLeuAspArgSerProTyrIleLeuLeuThreu 292
 Db 1006 AGTATGGAGAGAGTTTGAAGAGTCTGTGTCTGATCTCGATATGAGGAGATCTTGCCACC 1065
 Qy 293 LeuValIleLeuTyrGlyIleCysIleAsnLeuIleGlyValThrTyrPheSerGlnLeu 312
 Db 1066 CTGGTCGTGCTTATGGATATGAGCATTAACCTTGTGAGAGGTGATGAGTCAAAACTG 1125
 Qy 313 LysLeuGlnTyrProAsnMetAsnAspTyrSerGlnPheMetGlyAsnPheSerPheTyr 332
 Db 1126 AAGGCACAAATACCAACCCCTAATGATATCTTCATTCATGCGGAGATTCGACAGCT 1185
 Qy 333 ThrGlyValValSerValIleMetLeuPheValGlyIleAsnValIleGlyPhe 352
 Db 1186 ACTGGCAATACCAATTAATGATGATGTTGTTAGGAGAGTAAATCTTAGGAGAGTTT 1242
 Qy 353 GlyTyrPheThrGlyValLeuValThrProValMetValLeuLeuThrGlyIleValPhe 372
 Db 1243 GGTGGGAGTCCGAGCTTAATCAATCACCCCTGCACTGTCTTCATCTGAGATTGATTC 1302
 Qy 373 PheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrThr 392
 Db 1303 TTCTCACTGATTTTATTCGATGAGCCATGATCTCTTATGGCCACACTTGGAATGACA 1362
 Qy 393 ProLeuMetLeuAlaValIleValGlyValIleGlnAsnIleLeuSerIleSerThrLys 412
 Db 1363 CCTTTCCTCTGCGGTCTATGTTGGGCAATGAGAACATTTTCAGTAAGAGTCCGAG 1422
 Qy 413 TyrAlaLeuPheAspSerThrLysGlnMetAlaTyrIleProLeuAspGlnGlnLys 432
 Db 1423 TACAGTTGTTGACCTTGTGAGAAATGCAATCTCTTGTGCGAGACATGAG 1482
 Qy 433 ValIleGlyLysAlaAlaIleAspValIleAlaAlaIleAspPheGlyLysSerGlyValAla 452
 Db 1483 GTCAAGATTAAGAGGACGAATGATGTGTGCAACCATTTGGGAGAGTGTGAGGGCC 1542
 Qy 453 LeuIleGlnGlnLeuLeuValIleCysGlySerIleGlyAlaMetThrProTyrLeu 472
 Db 1543 TTGATTCAGAGATTCATTAATGATGATCTCTCCAAATTTACACCTTACTG 1602
 Qy 473 AlaValIleLeuPheIleAlaIleTyrLeuValSerAlaThrLysLeuAsnLys 492
 Db 1603 GAGGAATTTATTTGATTTGTTCTTCTGCTGTTGGTCTGTAAGTCCCTTGACTCA 1662
 Qy 493 LeuPheLeuAlaGlnSerAlaLeuLysGlnGlnValAlaGlnGlnAspSerAlaPro 512
 Db 1663 CAGTTC-----TCCCACTGGCAAGCAAGAGCTTGAGAGAAATGCTGAAA 1713
 Qy 513 AlaSerSer 515
 Db 1714 GCAAGAGC 1722
 RESULT 13
 AX505999 1875 bp DNA linear PAT 27-SEP-2002
 LOCUS Sequence 694 from Patent WO0216655.
 DEFINITION AX505999
 VERSION AX505999.1 GI:23387236
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 REFERENCE 1
 AUTHORS Harper, J.F., Kreps, J., Wang, X. and Zhu, T.
 TITLE Stress-regulated genes of plants, transgenic plants containing
 same, and methods of use
 JOURNAL Patent: WO 0216655-A 694 28-FEB-2002;
 The Scripps Research Institute (US); Syngenta Participations AG

Pred. No.: 6,61e-109 Length: 1875
 Score: 1333.00 Matches: 259
 Percent Similarity: 69.608 Conservative: 89
 Best Local Similarity: 51.808 Mismatches: 140
 Query Match: 50.684 Indels: 12
 DB: 6 Gaps: 3

US-09-869-433-2 (1-515) x AK651697 (1-1875)

QY 16 LeuTTPProIleHisThrHisgluLeuValysValLeuProMetPheLeuMetPhe 35
 DB 313 ATTTCGGGTGAGAGGTGCAACCTTGAAAGAAAGATTATCCCTTTAGATTGATGTTCTTT 372
 QY 36 CysIleThrPheAsnThrThrValLeuAspThrLysAspThrLysValLeuValLeu 55
 DB 373 TGTATCTTTCAATTACAAATTCAGAGGATCAAGAGATGTTCTTGTTGTTGAGCGG 432
 QY 56 ProGlySerGlyAlaGluAlaIleProPheIleLysPheTrpLeuValProCysAla 75
 DB 433 AAGAGAGTTCGTGAGATTATACCTTTCTTGAAACACTGGGTGAACTCTCTATAGCC 492
 QY 76 IleIlePheMetLeuIleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPhe 95
 DB 493 ATTTGGGTTATGCTCTCTACACTAAACTCCCAATGTTCTCTCCAGAGAGCTGTGTT 552
 QY 96 TyrAlaValGlyThrProPheLeuIlePhePheAlaLeuPheProThrValIleTyrPro 115
 DB 553 TACACTGTTATGTCCTTTCACTACATCTTGGGGCTTTGGTTTCTGTCATGATACCT 612
 QY 116 LeuArgAspValLeuHisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProPro 135
 DB 613 CTCAGAGATATATATACCCGAGAGCTCGCAGATACCTCTTACAAACCTCGGCCCA 672
 QY 136 GlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTyrValLeu 155
 DB 673 AGATTTCATGGCTTCAATTCATATATCCGAGATTGGAGTTCTGTTGTTATGTTATG 732
 QY 156 AlaGluLeuTrpGlySerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIle 175
 DB 733 GCTTGAGCTTGGGAGT 792
 QY 176 ThrLysIleHisGluAlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSer 195
 DB 793 ACACTGTGATGAGCAACCAAAATTCCTTCTTGTGAGACTTGAGCCCATGTGCA 852
 QY 196 LeuLeuAlaSerGlyValArgAlaIleValTyrAlaSerLysLeuArgAlaSerValSerGlu 215
 DB 853 CTGATTTTCTCAGGAAGAACCGTGAATCTCTTCACTTGAAGAAATCTTGTGCTCT 912
 QY 216 GlyValAspProTrpGlyLysLeuArgLeuLeuMetAlaMetThrIleValSerGly 235
 DB 913 GAGATTGAGCGCTGGCGAGTTTCGTGGAAGCCATGATGAGATTGTGTGGGAATGGGA 972
 QY 236 LeuValLeuMetAlaSerTyrTrpTrpIleAsnLysAsnValLeuThrAspProArgPhe 255
 DB 973 CTGGCATTTTGTCTCTCTATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1032
 QY 256 TyrAsnProGluGluMetGlnLysGlyLysGlyAlaLysProLysMetAsnMetLys 275
 DB 1033 AAGAACCAAGACAGAG-----AAACCAAGATGGAAGCATG 1068
 QY 276 AspSerPheLeuTyrLeuAspArgSerProTyrIleLeuLeuLeuThrLeuValIle 295
 DB 1069 GAACCTTGAGAGTCTTGTGTATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1128
 QY 296 AlaTyrGlyIleCysIleAsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGln 315
 DB 1129 GCATCGGATATAGATCATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1188
 QY 316 TyrProAsnMetAsnAspTyrSerGluPheMetGlyAsnPheSerPheTrpThrGlyVal 335
 DB 1189 TTCCTAGCCCAATGAGTACTCAGCATTTATGAGAGACTTCTCAACCTGACGCGGTGTT 1248

QY 336 ValSerValIleLeuMetLeuPheValGlyLysValIleArgLysPheGlyTrpLeu 355
 DB 1249 GCACATTCACAAATGATCTT---CTCAGCCAAATACGATTCATTAATGATGTTGGGA 1305
 QY 356 ThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeu 375
 DB 1306 GTAGCTGCAGAAATCACCCCAACTGTTCTGCTATGATGATGATGATGATGATGATGATG 1365
 QY 376 ValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMet 395
 DB 1366 ATATTGTTGGCGGCCCATTCGACCACTTGTGTGCAAGCTTGTGTATGACACCGCTACTT 1425
 QY 396 LeuAlaValAlaValGlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyrAlaLeu 415
 DB 1426 GCAGCTGTGTATGT 1485
 QY 416 PheAspSerThrLysGluMetAlaTyrIleProLeuAspGlnGluGlnLysValLysGly 435
 DB 1486 TTCGACCTTCGCAAGAAATGCGCTATATCCATTTGATGATGATGATGATGATGATGATG 1545
 QY 436 LysAlaAlaIleAspValAlaAlaAlaArgPheGlyLysSerGlyValAlaLeuIleGln 455
 DB 1546 AAAGCTGGATTTGACGTTGTGTGCAACCCATTAGGAAATAGAGGAGCTTTATATACAG 1605
 QY 456 GlnGlyLeuLeuValIleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValIle 475
 DB 1606 CAGTTCAATGATCTTATCTTGTGATCTACTACGAAATTCAGCGCTATCTGATATGATC 1665
 QY 476 LeuLeuPheIleIleAlaIleTyrPheValSerAlaThrLysLeuAsnLysLeuPheLeu 495
 DB 1666 TTGTTGTTATTTGTCACGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1722
 QY 496 AlaGlnSerAlaLeuLysGluGlnGluValAlaGlnIleAspSerAlaProAlaSerSer 515
 DB 1723 -----AACAGCTTGGCTGTGAAGAAAGCTTGAAGAAATGAGAGAGCTTCATCG 1776
 RESULT 15
 AF428316 2110 bp mRNA linear PLN 18-OCT-2001
 LOCUS Arabidopsis thaliana At1g80300/FS16_5 mRNA, complete cde.
 DEFINITION AF428316
 VERSION AF428316.1 GI:16226733
 KEYWORDS FLI CDNA.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 REFERENCE Arabidopsis thaliana; Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosoids; euroside II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 2110)
 CHEUK, R., CHEN, H., KIM, C.J., KOESEME, E., MEYERS, M.C., BANH, J., BOWSER, L., CARNINCI, P., DALE, J.M., GOLDSMITH, A.D., HAYASHIZAKI, Y., ISHIDA, J., JIANG, P.X., JONES, T., KAMIYA, A., KARLIN-NEUMANN, G., KAWAI, J., LAM, B., LEE, J.M., LIN, J., LIN, S.X., MIRANDA, M., NARUSAKA, M., NGUYEN, M., ONODERA, C.S., PALM, C.J., PHAM, P.K., QUACH, H.L., SAKURAI, T., SATOU, M., SEKI, M., SOUTHWICK, A., TANG, C.C., TORTUMLI, M., YAMADA, K., YAMAMURA, Y., YU, G., YU, S., SHINOZAKI, K., DAVIS, R.W., THEOLOGIS, A. and ECKER, J.R.
 Arabidopsis CDNA clones
 Unpublished
 2 (bases 1 to 2110)
 CHEUK, R., CHEN, H., KIM, C.J., KOESEME, E., MEYERS, M.C., BANH, J., BOWSER, L., CARNINCI, P., DALE, J.M., GOLDSMITH, A.D., HAYASHIZAKI, Y., ISHIDA, J., JIANG, P.X., JONES, T., KAMIYA, A., KARLIN-NEUMANN, G., KAWAI, J., LAM, B., LEE, J.M., LIN, J., LIN, S.X., MIRANDA, M., NARUSAKA, M., NGUYEN, M., ONODERA, C.S., PALM, C.J., PHAM, P.K., QUACH, H.L., SAKURAI, T., SATOU, M., SEKI, M., SOUTHWICK, A., TANG, C.C., TORTUMLI, M., YAMADA, K., YAMAMURA, Y., YU, G., YU, S., SHINOZAKI, K., DAVIS, R.W., THEOLOGIS, A. and ECKER, J.R.
 Direct Submission
 Submitted (05-OCT-2001) Salk Institute Genomic Analysis Laboratory (SIGAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

[illegible]

Search completed: November 25, 2003, 12:09:54
Job time : 6065 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:12:27 ; Search time 28 Seconds

(without alignments)
1768.817 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 2630

Sequence: 1 MTKTEKRPFKLRSFLMPH.....AQSLKQEVAAQEDSAPASS 515

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: PIR.76:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------|
| 1 | 2617 | 99.5 | 515 | 2 | E72089 |
| 2 | 2610 | 99.2 | 515 | 2 | E86534 |
| 3 | 2147 | 81.6 | 528 | 2 | C71561 |
| 4 | 2143 | 81.5 | 529 | 2 | C81774 |
| 5 | 1333 | 50.7 | 624 | 2 | E86834 |
| 6 | 1330.5 | 50.6 | 618 | 2 | G86288 |
| 7 | 1306 | 49.7 | 631 | 2 | T07420 |
| 8 | 1220.5 | 46.4 | 589 | 1 | S68205 |
| 9 | 1076 | 40.9 | 498 | 1 | A97710 |
| 10 | 1056 | 40.2 | 498 | 1 | J00026 |
| 11 | 1025.5 | 39.0 | 540 | 2 | C86567 |
| 12 | 1025.5 | 39.0 | 540 | 2 | B72056 |
| 13 | 976.5 | 37.1 | 501 | 2 | B97790 |
| 14 | 969.5 | 36.9 | 501 | 2 | B71707 |
| 15 | 956.5 | 36.4 | 540 | 2 | E71503 |
| 16 | 953 | 36.2 | 543 | 2 | F81665 |
| 17 | 907.5 | 34.5 | 499 | 2 | B97842 |
| 18 | 907.5 | 34.5 | 511 | 2 | B97783 |
| 19 | 880.5 | 33.5 | 512 | 2 | F71653 |
| 20 | 865 | 32.9 | 507 | 2 | B97765 |
| 21 | 827 | 31.4 | 507 | 2 | B97765 |
| 22 | 824 | 31.3 | 507 | 2 | B71695 |
| 23 | 229.5 | 8.7 | 441 | 2 | H82642 |
| 24 | 173.5 | 6.6 | 928 | 2 | F71541 |
| 25 | 161 | 6.1 | 918 | 2 | B81695 |
| 26 | 158.5 | 6.0 | 925 | 2 | A72096 |
| 27 | 157.5 | 6.0 | 925 | 2 | E81573 |
| 28 | 157.5 | 6.0 | 925 | 2 | E86527 |
| 29 | 151 | 5.7 | 1002 | 2 | AF2363 |

| | | | | | | |
|-----|-------|-----|------|---|--------|--|
| 30 | 145 | 5.5 | 540 | 2 | B82219 | transporter, BCCT |
| 31 | 144 | 5.5 | 660 | 2 | E81839 | probable transmembrane transporter, BCCT |
| 32 | 144 | 5.5 | 675 | 2 | E81101 | multidrug resistant |
| 33 | 140 | 5.3 | 474 | 2 | B83719 | cytochrome-c oxidase |
| 34 | 137 | 5.2 | 536 | 2 | T07946 | NADH2 dehydrogenase |
| 35 | 136.5 | 5.2 | 498 | 2 | T14236 | probable MFS trans |
| 36 | 135 | 5.1 | 441 | 2 | C83615 | oligosaccharyl transferase |
| 37 | 134 | 5.1 | 975 | 2 | E75212 | beta protein homol |
| 38 | 133 | 5.1 | 669 | 2 | D64137 | nitrite extrusion |
| 39 | 132.5 | 5.0 | 463 | 1 | GRECNK | nitrite extrusion |
| 40 | 131.5 | 5.0 | 463 | 2 | H90844 | nitrite extrusion |
| 41 | 131.5 | 5.0 | 463 | 2 | G85702 | nitrite extrusion |
| 42 | 131.5 | 5.0 | 567 | 2 | A99445 | permease, multidrug |
| 43 | 131 | 5.0 | 434 | 2 | AC2833 | hypothetical prote |
| 44 | 131 | 5.0 | 434 | 2 | G97610 | hypothetical prote |
| 45 | 130.5 | 5.0 | 976 | 2 | C71248 | hypothetical prote |
| 46 | 129.5 | 4.9 | 446 | 2 | A10401 | probable membrane |
| 47 | 129.5 | 4.9 | 581 | 2 | B90539 | hypothetical prote |
| 48 | 129 | 4.9 | 462 | 2 | S11431 | nitrite extrusion |
| 49 | 129 | 4.9 | 484 | 2 | T07782 | cytochrome-c oxidase |
| 50 | 128.5 | 4.9 | 435 | 2 | C89795 | hypothetical prote |
| 51 | 128.5 | 4.9 | 528 | 2 | S26948 | cytochrome-c oxidase |
| 52 | 128.5 | 4.9 | 665 | 2 | H87468 | ubiquinol oxidase |
| 53 | 128 | 4.9 | 462 | 2 | H90330 | amino acid transp |
| 54 | 127.5 | 4.8 | 505 | 2 | B88206 | protein F21D12.3 |
| 55 | 127 | 4.8 | 462 | 2 | H90887 | nitrite extrusion |
| 56 | 127 | 4.8 | 462 | 2 | A85730 | nitrite extrusion |
| 57 | 126.5 | 4.8 | 512 | 2 | G69670 | glycine betaine tr |
| 58 | 126.5 | 4.8 | 526 | 2 | S62763 | cytochrome-c oxidase |
| 59 | 126 | 4.8 | 548 | 2 | T51035 | hypothetical prote |
| 60 | 126 | 4.8 | 664 | 1 | B32243 | phosphotransferase |
| 61 | 125.5 | 4.8 | 641 | 2 | S44253 | raffinose carrier |
| 62 | 124 | 4.7 | 391 | 2 | D71676 | bicyclic protein re |
| 63 | 123.5 | 4.7 | 394 | 2 | S39739 | efflux protein hom |
| 64 | 123.5 | 4.7 | 531 | 2 | S78190 | cytochrome-c oxidase |
| 65 | 122.5 | 4.7 | 465 | 2 | AG0648 | nitrite extrusion |
| 66 | 122.5 | 4.7 | 512 | 2 | E83060 | iron (III)-transpo |
| 67 | 122.5 | 4.7 | 1139 | 2 | A10379 | probable potassium |
| 68 | 122 | 4.6 | 643 | 2 | E69373 | conserved hypotet |
| 69 | 121.5 | 4.6 | 894 | 2 | S45135 | probable membrane |
| 70 | 121 | 4.6 | 536 | 2 | S63651 | cytochrome-c oxidase |
| 71 | 120.5 | 4.6 | 472 | 2 | F90078 | hypothetical prote |
| 72 | 120.5 | 4.6 | 1118 | 2 | B83018 | conserved hypotet |
| 73 | 120 | 4.6 | 389 | 1 | C64923 | chloramphenicol re |
| 74 | 120 | 4.6 | 787 | 2 | H70374 | NADH2 dehydrogenase |
| 75 | 119.5 | 4.5 | 391 | 2 | AG0533 | probable drug effl |
| 76 | 119 | 4.5 | 389 | 2 | E90924 | probable transport |
| 77 | 119 | 4.5 | 389 | 2 | B85773 | probable transport |
| 78 | 119 | 4.5 | 493 | 2 | AG1773 | drug-export protei |
| 79 | 119 | 4.5 | 1131 | 2 | A97650 | hypothetical prote |
| 80 | 119 | 4.5 | 1131 | 2 | A82873 | two component sens |
| 81 | 118 | 4.5 | 462 | 2 | AC0672 | nitrite extrusion |
| 82 | 117.5 | 4.5 | 504 | 2 | E64118 | sodium/proline sym |
| 83 | 117.5 | 4.5 | 507 | 2 | AB1707 | glycine betaine tr |
| 84 | 117.5 | 4.5 | 512 | 2 | E89978 | high affinity prol |
| 85 | 117.5 | 4.5 | 744 | 2 | AD1382 | transport protein |
| 86 | 117 | 4.4 | 493 | 2 | AD1398 | drug-export protei |
| 87 | 117 | 4.4 | 684 | 2 | E97943 | Na+/H+ antiporter |
| 88 | 116.5 | 4.4 | 434 | 2 | H83227 | hypothetical prote |
| 89 | 116.5 | 4.4 | 522 | 2 | S25956 | cytochrome-c oxidase |
| 90 | 116.5 | 4.4 | 1126 | 2 | A96032 | probable two-compo |
| 91 | 116 | 4.4 | 472 | 2 | T34748 | transmembrane tran |
| 92 | 116 | 4.4 | 475 | 2 | A70330 | nitrate transp |
| 93 | 116 | 4.4 | 635 | 1 | A64162 | cytochrome c-type |
| 94 | 116 | 4.4 | 684 | 2 | B96308 | conserved hypotet |
| 95 | 115.5 | 4.4 | 397 | 2 | D65066 | hypothetical prote |
| 96 | 115.5 | 4.4 | 1050 | 2 | G70396 | cation efflux syst |
| 97 | 115 | 4.4 | 406 | 2 | T43120 | conserved hypotet |
| 98 | 115 | 4.4 | 487 | 2 | A40956 | probable membrane |
| 99 | 115 | 4.4 | 521 | 2 | A32431 | cytochrome-c oxidase |
| 100 | 115 | 4.4 | 541 | 2 | A48327 | cytochrome-c oxidase |

ALIGNMENTS

RESULT 1

ADP, ATP carrier protein CP0408 [imported] - Chlamydia pneumoniae (strain CML029 at E72089
 C|Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C|Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C|Accession: E72089; D81580
 R|Kalmann, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999.
 A|Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A|Reference number: A72000; MUID:99206606; PMID:10192388
 A|Accession: E72089
 A|Molecule type: DNA
 A|Residues: 1-515 <ARN>
 A|Cross-references: GB:AE001619; GB:AE001363; NID:94376620; PIDN:AA018495.1; PID:9437662
 A|Experimental source: strain CML029
 R|Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A|Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A|Reference number: A81500; MUID:20150255; PMID:10684935
 A|Accession: D81580
 A|Molecule type: DNA
 A|Residues: 1-515 <REA>
 A|Cross-references: GB:AE002202; GB:AE002161; NID:97189324; PIDN:AA38252.1; PID:9718933
 A|Experimental source: strain AR39, HL cells
 C|Genetics:
 A|Gene: act 1; CP0408
 C|Superfamily: rickettsial-type ATP/ADP translocase

Query Match 99.5%; Score 2617; DB 2; Length 515;
 Best Local Similarity 99.6%; Pred. No. 8.7e-193;
 Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 MTKEEKPFGLRSPFLPIHTEHLKVLPMFLMFCITFNTYTVLRDTKDTLIVGAPSGA 60
DB 1 MTKEEKPFGLRSPFLPIHTEHLKVLPMFLMFCITFNTYTVLRDTKDTLIVGAPSGA 60
QY 61 EAIPFIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPELLIFALPFTIYPLRDVL 120
DB 61 EAIPFIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPELLIFALPFTIYPLRDVL 120
QY 121 HPTFADRLQAILPPGLIGVAILRNWTFAPFYLAELMGVMSLMFWGFANETIKHE 180
DB 121 HPTFADRLQAILPPGLIGVAILRNWTFAPFYLAELMGVMSLMFWGFANETIKHE 180
QY 181 AKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLMAAMTIVSGLVMA 240
DB 181 AKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLMAAMTIVSGLVMA 240
QY 241 SYMWINKNVLTDPREFYNEEMQKKGAKPKNNKDSFLYLDSPYILLTLVIAVIGIC 300
DB 241 SYMWINKNVLTDPREFYNEEMQKKGAKPKNNKDSFLYLDSPYILLTLVIAVIGIC 300
QY 301 INLEEVTKSQKLQYPMNDYSEFMGNFSFWTGVSVLIMLVGNAVIRKFGWLTGALV 360
DB 301 INLEEVTKSQKLQYPMNDYSEFMGNFSFWTGVSVLIMLVGNAVIRKFGWLTGALV 360
QY 361 TPVAVLLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTK 420
DB 361 TPVAVLLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTK 420
QY 421 EMAYIPLDQEQKVKGAIDVVAARFGKSGALLIQGLLVICSGIGAMTPYLAIVLLFII 480
DB 421 EMAYIPLDQEQKVKGAIDVVAARFGKSGALLIQGLLVICSGIGAMTPYLAIVLLFII 480
QY 481 AIWLVSAATKLNKFLAOSALKQEVVAOEDSAPASS 515
DB 481 AIWLVSAATKLNKFLAOSALKQEVVAOEDSAPASS 515

```

RESULT 2

BE6534
 ADP/ATP translocase [imported] - Chlamydia pneumoniae (strain J138)
 C|Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C|Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C|Accession: BE6534
 R|Shirai, M.; Hirkawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ishiura, H.; Nucleic Acids Res. 29, 2311-2314, 2000
 A|Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
 A|Reference number: A86491; MUID:20330349; PMID:10871362
 A|Accession: BE6534
 A|Status: preliminary
 A|Molecule type: DNA
 A|Residues: 1-515 <STO>
 A|Cross-references: GB:BA000008; NID:98978723; PIDN:BA098559.1; GSPDB:GN00142
 A|Experimental source: strain J138
 C|Genetics:
 A|Gene: act 1
 C|Superfamily: rickettsial-type ATP/ADP translocase

Query Match 99.2%; Score 2610; DB 2; Length 515;
 Best Local Similarity 99.4%; Pred. No. 3e-192;
 Matches 512; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 MTKEEKPFGLRSPFLPIHTEHLKVLPMFLMFCITFNTYTVLRDTKDTLIVGAPSGA 60
DB 1 MTKEEKPFGLRSPFLPIHTEHLKVLPMFLMFCITFNTYTVLRDTKDTLIVGAPSGA 60
QY 61 EAIPFIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPELLIFALPFTIYPLRDVL 120
DB 61 EAIPFIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPELLIFALPFTIYPLRDVL 120
QY 121 HPTFADRLQAILPPGLIGVAILRNWTFAPFYLAELMGVMSLMFWGFANETIKHE 180
DB 121 HPTFADRLQAILPPGLIGVAILRNWTFAPFYLAELMGVMSLMFWGFANETIKHE 180
QY 181 AKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLMAAMTIVSGLVMA 240
DB 181 AKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLMAAMTIVSGLVMA 240
QY 241 SYMWINKNVLTDPREFYNEEMQKKGAKPKNNKDSFLYLDSPYILLTLVIAVIGIC 300
DB 241 SYMWINKNVLTDPREFYNEEMQKKGAKPKNNKDSFLYLDSPYILLTLVIAVIGIC 300
QY 301 INLEEVTKSQKLQYPMNDYSEFMGNFSFWTGVSVLIMLVGNAVIRKFGWLTGALV 360
DB 301 INLEEVTKSQKLQYPMNDYSEFMGNFSFWTGVSVLIMLVGNAVIRKFGWLTGALV 360
QY 361 TPVAVLLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTK 420
DB 361 TPVAVLLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTK 420
QY 421 EMAYIPLDQEQKVKGAIDVVAARFGKSGALLIQGLLVICSGIGAMTPYLAIVLLFII 480
DB 421 EMAYIPLDQEQKVKGAIDVVAARFGKSGALLIQGLLVICSGIGAMTPYLAIVLLFII 480
QY 481 AIWLVSAATKLNKFLAOSALKQEVVAOEDSAPASS 515
DB 481 AIWLVSAATKLNKFLAOSALKQEVVAOEDSAPASS 515

```

RESULT 3

C71561
 Probable adp/act translocase - Chlamydia trachomatis (serotype D, strain WJ3/Cx)
 C|Species: Chlamydia trachomatis
 C|Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 20-Aug-1999
 C|Accession: C71561
 R|Stephens, R.S.; Kalmann, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
 A|Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia traci
 A|Reference number: A71570; MUID:9900809; PMID:9784136
 A|Accession: C71561
 A|Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-528 <ARN>
 A:Cross-references: GB:AE001281; GB:AE001273; NID:g3328454; PIDN:AA067656.1; PID:g332845
 A:Experimental source: serotype D, strain UM-3/CX
 C:Genetics:
 A:Gene: CT065
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 81.6%; Score 2147; DB 2; Length 528;
 Best Local Similarity 78.1%; Pred. No. 8.8e-157;
 Matches 410; Conservative 52; Mismatches 49; Indels 14; Gaps 3;

```

Qy 1 MTKTEKPFGLKRSFLMPLIHTHELKVLPMFLMFCITFNNTYVLRDRTDILVAPSGA 60
Db 1 MTOAEKPFGLKRSFLMPLIHTHELKVLPMFLMFCISFNNTYLRDRTDILVAPSGA 60
Qy 61 EAIPIKFWLVPCAIIFMLIVAKLSNLSKQALFYAVGTPFLIFPALFTVIYPLRDVL 120
Db 61 EAIPIKFWLVPCAIIFMLIVAKLSNLSKQALFYAVGTPFLIFPALFTVIYPCRHIL 120
Qy 121 HPTDFADTLQAIIPSGLGLVALIRNMTFAFYVLAELMGSVMSLMFPGFANEITKISE 180
Db 121 HPTDFADTLQAIIPSGLGLVALIRNMTFAFYVLAELMGSVMSLMFPGFANEITKISE 180
Qy 181 AKRFYALFGIGANISLASGRAIVWASKRASVSEGVDPMGISRLIMAMTIYGLVMA 240
Db 181 AKRFYALFGIGANISLASGRAIVWASKRASVSEGVDPMGISRLIMAMTIYGLVMA 240
Qy 241 SYWMINKVLTDPREFYNEEMQKKGAKPKMNKDSFLYLDSPYLLTLTLVAVGIC 300
Db 241 SYWMINKVLTDPREFYNEEMQKKGAKPKMNKDSFLYLDSPYLLTLTLVAVGIC 300
Qy 301 INLEVTWKSQOLKQYNNNDYSEFNGNSFWTGVSVLIMLFGVGNVIRKFGMLTGLV 360
Db 301 INLEVTWKSQOLKQYNNNDYSEFNGNSFWTGVSVLIMLFGVGNVIRKFGMLTGLV 360
Qy 361 TPVWVLTGAVFPAVLVFRNQSGLVAMFGTTPMLAVVGAIONISKSTKVALPSTK 420
Db 361 TPVWVLTGAVFPAVLVFRNQSGLVAMFGTTPMLAVVGAIONISKSTKVALPSTK 420
Qy 421 EMAYIPLDQEQKXGKAIDVVAARFGSGGALIQGLILVCGSIGAMTPYLAVALFI 480
Db 421 EMAYIPLDQEQKXGKAIDVVAARFGSGGALIQGLILVCGSIGAMTPYLAVALFI 480
Qy 481 AIMVLSATKINKLFLAOSALKQEVAVQ-----EBSAPA 513
Db 481 AIMVLSATKINKLFLAOSALKQEVAVQ-----EBSAPA 513
Qy 513 MVLTSATKINKLFLAOSALKQEVAVQ-----EBSAPA 513
Db 513 MVLTSATKINKLFLAOSALKQEVAVQ-----EBSAPA 513

```

RESULT 4

ADP: ATP carrier protein TC035 [imported] - Chlamydia muridarum (strain Nigg)
 C/Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
 C/Accession: C81714
 R/Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, W.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A/Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A:Reference number: AB1500; MUID:20150255; PMID:10684935
 A/Accession: C81714
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-529 <TGT>
 A:Cross-references: GB:AE002301; GB:AE002160; NID:g7190372; PIDN:AAF39198.1; PID:g719037
 A:Experimental source: strain Nigg (Mopn)
 C:Genetics:
 A:Gene: TC035
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 81.5%; Score 2143; DB 2; Length 529;
 Best Local Similarity 78.5%; Pred. No. 1.8e-156;
 Matches 408; Conservative 55; Mismatches 49; Indels 8; Gaps 3;

```

Qy 1 MTKTEKPFGLKRSFLMPLIHTHELKVLPMFLMFCITFNNTYVLRDRTDILVAPSGA 60
Db 1 MTOAEKPFGLKRSFLMPLIHTHELKVLPMFLMFCISFNNTYLRDRTDILVAPSGA 60
Qy 61 EAIPIKFWLVPCAIIFMLIVAKLSNLSKQALFYAVGTPFLIFPALFTVIYPLRDVL 120
Db 61 EAIPIKFWLVPCAIIFMLIVAKLSNLSKQALFYAVGTPFLIFPALFTVIYPCRHIL 120
Qy 121 HPTDFADTLQAIIPSGLGLVALIRNMTFAFYVLAELMGSVMSLMFPGFANEITKISE 180
Db 121 HPTDFADTLQAIIPSGLGLVALIRNMTFAFYVLAELMGSVMSLMFPGFANEITKISE 180
Qy 181 AKRFYALFGIGANISLASGRAIVWASKRASVSEGVDPMGISRLIMAMTIYGLVMA 240
Db 181 AKRFYALFGIGANISLASGRAIVWASKRASVSEGVDPMGISRLIMAMTIYGLVMA 240
Qy 241 SYWMINKVLTDPREFYNEEMQKKGAKPKMNKDSFLYLDSPYLLTLTLVAVGIC 300
Db 241 SYWMINKVLTDPREFYNEEMQKKGAKPKMNKDSFLYLDSPYLLTLTLVAVGIC 300
Qy 301 INLEVTWKSQOLKQYNNNDYSEFNGNSFWTGVSVLIMLFGVGNVIRKFGMLTGLV 360
Db 301 INLEVTWKSQOLKQYNNNDYSEFNGNSFWTGVSVLIMLFGVGNVIRKFGMLTGLV 360
Qy 361 TPVWVLTGAVFPAVLVFRNQSGLVAMFGTTPMLAVVGAIONISKSTKVALPSTK 420
Db 361 TPVWVLTGAVFPAVLVFRNQSGLVAMFGTTPMLAVVGAIONISKSTKVALPSTK 420
Qy 421 EMAYIPLDQEQKXGKAIDVVAARFGSGGALIQGLILVCGSIGAMTPYLAVALFI 480
Db 421 EMAYIPLDQEQKXGKAIDVVAARFGSGGALIQGLILVCGSIGAMTPYLAVALFI 480
Qy 481 AIMVLSATKINKLFLAOSALKQEVAVQ-----AEDSAPAS 514
Db 481 AIMVLSATKINKLFLAOSALKQEVAVQ-----AEDSAPAS 514
Qy 514 MVLTSATKINKLFLAOSALKQEVAVQ-----AEDSAPAS 514
Db 514 MVLTSATKINKLFLAOSALKQEVAVQ-----AEDSAPAS 514

```

RESULT 5

adenine nucleotide translocase, 19474-21800 [imported] - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
 C/Accession: E96834
 R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizat, L.
 Nature 408, 816-820, 2000
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luroe, J.S.; Maiti, R.; Marzilli, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, J. ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: AB6141; MUID:21016719; PMID:11130712
 A/Accession: E96834
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-624 <STO>
 A:Cross-references: GB:AE005173; NID:g6751705; PIDN:AAF27687.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F516.5
 A:Map position: 1
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 50.7%; Score 1333; DB 2; Length 624;
 Best Local Similarity 51.8%; Pred. No. 2.3e-94;
 Matches 259; Conservative 89; Mismatches 140; Indels 12; Gaps 3;

```

Oy 76 IIFMLIYAKLSNIIISKQALFYAVGTPFLIPFALPFTVYLPRLDVLHPTFADRLQALP 135
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 165 IGFMLIYAKLSNIIISKQALFYAVGTPFLIPFALPFTVYLPRLDVLHPTFADRLQALP 224
Oy 136 GLIGLVALIRNMTFAFYVLAELMGSVLIMFMGFANEITKIHAKRPYALFGIGANIS 195
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 225 RFMGPIALIRIWSFCLFYVMAELMGSVVSVLFWGFANQITTVDEAKKFFLFGIGANVA 284
Oy 196 LLAGRAIIVASKIRASVSEGVDPWGISRLILMAMTIVSGVLVMAVYMWINKVLTDP 255
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 285 LIFGGRITKYFSPNIRKNLPGVDGVAVSLKGMMSIVGMGLAICLLYVWVRYPPLPTRS 344
Oy 256 YNPSEMOKGKGAAPKRNKMSFLYLDSPYIILLTLVIAVGINILEVTWKSQKQ 315
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 345 KKKKE-----KPKGMTESLKFVSSPYIRDLATLVAYGISINLVEVTWKSQKQ 396
Oy 316 YPNNDYSEFMNGSPFWTGVSVLIMLVGAGNVIRKFGMLTGAIVTPVWVLTGIVFPAL 375
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 397 FSPNEYSAFMGDSTCTGATFTMML-LSQYFVKYGMVAAKITPTVLLTGVAFPSL 455
Oy 376 VIPRNOASGLVAMGTTPLMLAVVGAIONLSTKTYALPDSKEMAYIPLDOQRYK 435
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 456 ILFGPPAPLVAKIGMTPPLAAVVGALQNTFSKAKYSLFDPCKEMAYIPLDPTVK 515
Oy 436 KKAIDVVAARFGKSGGALLIOGGLLVIGSIGAMPYLAIVILFIAIMLVSATKLKFL 495
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 516 KKAIDVVCNPLGKSGGALLIOGGLLVIGSIGAMPYLAIVILFIAIMLVSATKLKFL 574
Oy 496 AQSALKEQVAQEDSAPAS 515
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 575 --NSLRSEELEREMERASS 592
    |||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 6

```

G86288
probable adenine nucleotide translocase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
C:Accession: G86288
R:Thellogis, A.; Becker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
  Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
  ansen, N.F.; Hughes, B.; Hutzar, L.
  Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
  C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marzilli,
  Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
  ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G86288
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-618 <STD>
A:Cross-references: GB:AE005172; NID:G8072388; PIDN:AAF71976.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
C:Superfamily: rickettsial-type ATP/ADP translocase

```

```

Query Match 50.6%; Score 1330.5; DB 2; Length 618;
Best Local Similarity 50.9%; Pred. No. 3,5e-94;
Matches 260; Conservative 92; Mismatches 136; Indels 23; Gaps 5;

Oy 16 LMPHITHELKLVLPMLFPCITENYVLRDTKTLIVGAPGSAEAIPIKFWLVPCALIFMLIYA 75
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 102 IFGEVETTLKKIVPGMLFPCILENYITLRDTKDLVLTAKSSAEIIPFLKTVNVLPM 161
Oy 76 IIFMLIYAKLSNIIISKQALFYAVGTPFLIPFALPFTVYLPRLDVLHPTFADRLQALP 135
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 162 IGFMLIYAKLSNIIISKQALFYAVGTPFLIPFALPFTVYLPRLDVLHPTFADRLQALP 221
Oy 136 GLIGLVALIRNMTFAFYVLAELMGSVLIMFMGFANEITKIHAKRPYALFGIGANIS 195
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 222 RFMGPIALIRIWSFCLFYVMAELMGSVVSVLFWGFANQITTVDEAKKFFLFGIGANVA 281
Oy 196 LLAGRAIIVASKIRASVSEGVDPWGISRLILMAMTIVSGVLVMAVYMWINKVLTDP 255
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 282 LIFGGRITKYFSPNIRKNLPGVDGVAVSLKGMMSIVGMGLAICLLYVWVRYPPLPTRS 344
Oy 256 YNPSEMOKGKGAAPKRNKMSFLYLDSPYIILLTLVIAVGINILEVTWKSQKQ 315
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 345 KKKKE-----KPKGMTESLKFVSSPYIRDLATLVAYGISINLVEVTWKSQKQ 396
Oy 316 YPNNDYSEFMNGSPFWTGVSVLIMLVGAGNVIRKFGMLTGAIVTPVWVLTGIVFPAL 375
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 397 FSPNEYSAFMGDSTCTGATFTMML-LSQYFVKYGMVAAKITPTVLLTGVAFPSL 455
Oy 376 VIPRNOASGLVAMGTTPLMLAVVGAIONLSTKTYALPDSKEMAYIPLDOQRYK 435
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 456 ILFGPPAPLVAKIGMTPPLAAVVGALQNTFSKAKYSLFDPCKEMAYIPLDPTVK 515
Oy 436 KKAIDVVAARFGKSGGALLIOGGLLVIGSIGAMPYLAIVILFIAIMLVSATKLKFL 491
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 516 KKAIDVVCNPLGKSGGALLIOGGLLVIGSIGAMPYLAIVILFIAIMLVSATKLKFL 572
Oy 492 KLFPAQSLKQ-----VAQEDSAPA 513
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 573 TLMSEELEREMERASSVKIPVSGED-APS 602
    |||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 7

```

T07420
ATP/ADP-transporter, chloroplast - potato
C:Species: Solanum tuberosum (potato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 20-Jun-2000
C:Accession: T07420
R:Tjaden, J.; Mohlmann, T.; Kampfenkel, K.; Henrichs, G.; Neuhaus, H.E.
  Plant J. 16, 531-540, 1998
A:Title: Altered plastidic ATP/ADP-transporter activity influences potato (Solanum tuber
  A:Reference number: Z16025
A:Accession: T07420
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-631 <TUA>
A:Cross-references: EMBL:Y10821; PIDN:CAA71785.1
A:Experimental source: cv. Desire
C:Genetics:
A:Genome: nuclear
C:Superfamily: rickettsial-type ATP/ADP translocase
C:Keywords: chloroplast; membrane protein

```

```

Query Match 49.7%; Score 1306; DB 2; Length 631;
Best Local Similarity 52.1%; Pred. No. 2,7e-92;
Matches 256; Conservative 87; Mismatches 136; Indels 12; Gaps 3;

Oy 24 LKRVLPMLFPCITENYVLRDTKTLIVGAPGSAEAIPIKFWLVPCALIFMLIYA 83
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 111 LKRIIPGAMFPCILFNYITLRDTKDLVLTAKSSAEIIPFLKTVNVLPMALIGFMLIYT 170
Oy 84 KLSNIIISKQALFYAVGTPFLIPFALPFTVYLPRLDVLHPTFADRLQALPGLGLV 143
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 171 KLANVLSKEALFYVILPFAFGAFGVLYPSNHYHPTAFADKLTNTGPRFLGPAL 230
Oy 204 VMAKIRASVSEGVDPWGISRLILMAMTIVSGVLVMAVYMWINKVLTDPFYNPEMOK 263
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 291 KYFSSLRSSIGPVDGVAISLKGMSIVVMGALICFPYVWVRYPPLPTRS-----S 342
Oy 264 GKKAAPKRNKMSFLYLDSPYIILLTLVIAVGINILEVTWKSQKQYNNNDYS 323
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 343 KKKKVKNNMTMSLKFVSSPYIRDLATLVAAIGISINLVEVTWKSQKQAFSPSEYS 402
Oy 324 EFMNGSPFWTGVSVLIMLVGAGNVIRKFGMLTGAIVTPVWVLTGIVFPALVIPRNOAS 383
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db      403 SEMGPFSTATGATATFMML-LSQWIFDKXGMAAKITPTVLLTGVSFSLLEGAPLA 461
Qy      384 GLVAFPGTTPMLAVVGAIONILSKSTYALPDSTKEMAYIPLDQEQVYKGAIDVYA 443
Db      462 PTLAKFGMTPLAAVVGAMONIFSKSAKYSLEDPCKEMAYIPLEDDTQVYKKAIDVYC 521
Qy      444 ARFGKSGGALLIOGGLVIGSIGAMTPYLAIVLFIATWLSATKTLKPLAQSALKRQ 503
Db      522 NPLGSGGALLIOGFMILTLFTGSLASTPYLGVLLVIVLAMLGAASLDSQF---TQLRQ 578
Qy      504 EVAQEDSAPAS 514
Db      579 EDLEKEMERAS 589

```

RESULT 8

S68205
ATP/ADP translocase ATP1 precursor - Arabidopsis thaliana
N:Alternate names: adenine nucleotide translocase; ATP/ADP translocator protein
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 22-Jun-1999
C:Accession: S68205
R:Kampfenkel, K.; Moehlmann, T.; Batz, O.; Van Montagu, M.; Inze, D.; Neuhaus, H.E.
FBS Lett. 374, 351-355, 1995
A:Title: Molecular characterization of an Arabidopsis thaliana cDNA encoding a novel put
A:Reference number: S68205; MUID:96069943; PMID:7589569
A:Accession: S68205
A:Molecule type: mRNA
A:Residues: 1-589 <KAM>
A:Cross-references: EMBL:Z49227; NID:g1051108; PIDN:CAA89201.1; PID:g1051109
C:Genetics:
A:Gene: ATP1
A:Genome: nuclear
C:Superfamily: rickettsial-type ATP/ADP translocase
C:Keywords: Chloroplast; transmembrane protein
P:1-100/Domain: translt peptide (chloroplast) #status predicted <TNP>
P:101-589/Product: ATP/ADP translocase ATP1 #status predicted <MNT>

```

Query Match      46.4%; Score 1220.5; DB 1; Length 589;
Best Local Similarity 51.7%; Pred. No. 8.8e-86;
Matches 252; Conservative 76; Mismatches 144; Indels 15; Gaps 4;

Qy      16 LMPITHLKKVLPMLPFECITFNTYVLRDRTDLYVAGPSGAELPFIKMLVPCA 75
Db      104 IFGVAVATLKTIIPGLMEFCILFNTYILRDKDVLAVYAKGSAEIIPLFTWNLPA 163
Qy      76 IIFMLIVAKLSNLSKQALFYAVGTFPLIFALFPTVIYPLRDVLPHTFEADRLQALIPP 135
Db      164 IGFMLITKLSNVLSKQALFYVIVPFIIFGFGFWYPLSNYIHPEALDKLTTTGF 223
Qy      136 GLLGLVAILRNWTFAPFYVLAELWGSVMSLWFGFANEITKHEAKRFYALFGIGANIS 195
Db      224 RFMGFIILIRIWSFCIFYMAELMGSVVSVLFMGFANDITVDAAKFPYLFIGANVA 283
Qy      196 LLSGRALVMSKLRASVSEGVDPWGISRLMLAMTIYSGVLMSYTMINKVLTDRF 255
Db      284 LIFSRTVYVFNLRKNKNGVD--GSFESHDEHCGGNGTRI CUSTIGSNVYVLPRTS 341
Qy      256 YNPEEMOKGKAKPKPMNKDSFLYIDRSPIILLTLVIAYGICINLIEVTKSQKLQ 315
Db      342 KKXKE-----KKMGMSLKLFLVSPYIRDLATLVVAYGISINLVEVTKSKLQ 393
Qy      316 YPNNDYSEFMGNESFMTGVSVLIMLFYGVGNVIRKFGMLTGALTVPVWVLLTGIVPAL 375
Db      394 FSPMEYSAFMGAFSTCTGATFTMML-LSQYVFNKYGWGAAKITPTVLLTGVAFFSL 452
Qy      376 VIFRNQASGLVAFGTTPMLAVVGAIONILSKSTYALPDSTKEMAYIPLDQEQVYK 435
Db      453 ILFGPFPALVAKLGMTPLAAVVGALONISKSAKSLPCKEMAYIPLEDDTQVYK 512
Qy      436 KKAIDVVARFSGSGGALLIOGGLVIGSIGAMTPYLAIVLFIATWLSATKTL-----N 491

```

```

Db      513 KKAIDVVCNPLKSGGALLIOGFMILSPGSLANSTPYLGMILLVITWAMLAASLEGQFN 572
Qy      492 KLFPLAQS 498
Db      573 SLRLKKS 579

```

RESULT 9

A97710
adp/ATP carrier protein [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: A97710
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barde, V.; Samson, D.; R
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: A97710
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-498 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL02619.1; PID:g15619119; GSPDB:GN00173
C:Genetics:
A:Gene: tlc1
C:Superfamily: rickettsial-type ATP/ADP translocase

```

Query Match      40.9%; Score 1076; DB 2; Length 498;
Best Local Similarity 43.6%; Pred. No. 8.6e-75;
Matches 213; Conservative 103; Mismatches 153; Indels 20; Gaps 7;

Qy      11 KLRSLFPIHHEKLVLPMLPFECITFNTYVLRDRTDLYVAGPSGAELPFIKML 70
Db      12 ELKSVIPIERENKFLPMAFMFCILNLTSLKSDGEV--TDIGAELISFLKTYI 69
Qy      71 VPCAILFMLIVAKLSNLSKQALFYAVGTFPLIFALFPTVIYPLRDVLPHTFEADRLQ 130
Db      70 VLPASVIMAVIVYKCDILKQENVYVITSFPLGVFALFAPVLYPPLVHPDEPTISW 129
Qy      131 AILPPLGLVAILRNWTFAPFYVLAELWGSVMSLWFGFANEITKHEAKRFYALFGI 190
Db      130 SVAYPNVMPFRIRIVKMSFASFTYMAELMGTMISLTFWOFANQITKTDEAKRFYSMGL 189
Qy      191 GANISLASGRAIYVASTLRASVSEGVDPWGISRL-----LMAMTIYSGVLMSYTMINK 246
Db      190 LANLALPYTSVITIGCLHEKTIQI-----VAELKRVLPFLVMTTSFLVITVYRWKN 241
Qy      247 KNVLTDPFRFYNPEEMOKGKAKPKPMNKDSFLYIDRSPIILLTLVIAYGICINLIEV 306
Db      242 KNVLTDPRLYDPAVYK--EKAKAKMSLIDSPKMLFTSKYGYIALMLIAYGVSVNLVGG 299
Qy      307 TWKSQKLQYPNNDYSEFMGNESFMTGVSVLIMLFYGVGNVIRKFGMLTGALTVPVWV 366
Db      300 VKMSKVKELVPTKEAYTITMGKFPQYQGVVAIAFWL--IGSNILRRVSWLTAMITPLMML 358
Qy      367 LTGIYFELVLPFRNG--ASGLVAMGTTPMLAVVGAIONILSKSTYALPDSTKEMAYI 425
Db      359 ITGAAPFAFIPFDSTYIAHLTGILASGPLAAVWGMQONVLSKGVKSLDPATONMYI 418
Qy      426 PLDQEQVYKGAIDVVARFSGSGGALLIOGGLVIGSIGAM--TPYLAIVLFIATW 483
Db      419 PLDDRLKRGQAAVAVIGRFGKSGGALLIOGTFTPLPAPGVVATPFAISFPIVILW 478
Qy      484 LVSATKLNK 492
Db      479 IYAVKGLNK 487

RESULT 10
U00026
ATP/ADP translocase tlc1 - Rickettsia prowazekii
N:Alternate names: Adp/ATP carrier protein tlc1; RP053
C:Species: Rickettsia prowazekii
C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 03-Nov-2000

```

C:Accession: J00026; E71713
R:Williamson, L.R.; Plano, G.V.; Winkler, H.H.; Krause, D.C.; Wood, D.O.
Gene 80, 269-278, 1989
A:Title: Nucleotide sequence of the *Rickettsia prowazekii* ATP/ADP translocase-encoding
A:Reference number: J00026; MUID:90060776; PMID:2555259
A:Accession: J00026
A:Molecule type: DNA
A:Residues: 1-498 <NT>
A:Cross-references: GB:M2816; NID:9152469; PIDN:AAA6382.1; PID:9152470
R:Anderson, S.G.E.; Zomrodipou, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U
Nature 396, 133-140, 1998
A:Title: The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:9823893
A:Accession: E71713
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-498 <AND>
A:Cross-references: GB:AJ235270; GB:AJ235269; NID:93860572; PIDN:CAA14524.1; PID:9386062
A:Experimental source: strain Madrid E
A:Comment: This ATP/ADP translocase shares no detectable amino acid sequence homologies
C:Genetics:
A:Gene: tlc1; RP053
C:Function:
A:Description: exchanges ADP in the rickettsial cell with ATP in the host cell
C:Superfamily: rickettsial-type ATP/ADP translocase
C:Keywords: transmembrane protein
F:28-45/Domain: transmembrane #status predicted <TR01>
F:68-84/Domain: transmembrane #status predicted <TR02>
F:93-115/Domain: transmembrane #status predicted <TR03>
F:183-206/Domain: transmembrane #status predicted <TR04>
F:219-237/Domain: transmembrane #status predicted <TR05>
F:279-297/Domain: transmembrane #status predicted <TR06>
F:346-370/Domain: transmembrane #status predicted <TR07>
F:380-399/Domain: transmembrane #status predicted <TR08>
F:445-461/Domain: transmembrane #status predicted <TR09>
F:466-482/Domain: transmembrane #status predicted <TR10>

Query Match 40.2%; Score 1056; DB 1; Length 498;
Best Local Similarity 41.6%; Pred. No. 2.9e-73;
Matches 207; Conservative 106; Mismatches 165; Indels 20; Gaps 7;

Qy 2 TKEEPEFGKLRSPFPIHTEHLEKVLPMFPCITFNYTVLRDTOTLIIGARGSGAE 61
Db 3 TKSSENVSELRKIIWPIEQYENKFLPLAFMFCILNYSLTRSIKSGFV--IDIGRE 60

Qy 62 AIPPIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPLIFPALTFTVYIPRLDVH 121
Db 61 SISFLKTYIVLPSAVIAMIIYVKLCIILQENVFYVITSFPLGYALFAFVLYPYDILVH 120

Qy 122 PTEFADRLQALIPGILGLVAIILRMWTFPAFVLAELMGSVMLSLMFGFANEIKIHEA 181
Db 121 POKHTESISLAVYPMFKFKIKIVGRKSPFYTIALMGTWMLSLFVQFAQITKIAEA 180

Qy 182 KRFVALFGIGANISILASGRAIYWASKLRASVEGDVPGISLR-----LMAITVSGIV 237
Db 181 KRFYMPFGILNIALPYSVITGYLHETQI-----VAHLKFPVLEFIMTSSGL 232

Qy 238 LMAIYMWINKNVLTPRFNYPEEMOKGKAGPRKNNKDSFLYLDSPYIILLITLITAY 297
Db 233 ILITRWNNKVNLTDPRLVDPALVK--EKTKAKLSFISLKMIFTSKVVGYIALITAY 290

Qy 298 GICILILEVTWSQKLQYPMNDYSEFNGNFSFMTGVVSVLIMFVGGNVIRKGMILTG 357
Db 291 GVSVALVGVGWSKVKELPTREAYTIVGQFQGWALAFML--IGSNILRKVSWLTA 349

Qy 358 ALVTPVWVLLTGVFPALVIFRNO--ASGLVAMFGTTPMLAVVGAIONILSKSTRYALF 416
Db 350 AMITPLMFPITGAFFSPFSDSVIAMLITGLASSPLTILAMIGMIVLSKGYKYSIF 409

Qy 417 DSTKEMAYIPLDQOKVKGKAAIDVVAARFGSGGALLIQGLLVI CGSIGAM--TPYLAV 474
Db 410 DATKMAYIPLDKDLRVKGAAVEVIGRLGKSGGALLIGSTFFILFPVGFIEATPYPAS 469

Qy 475 ILFFIATMVSATKANK 492
Db 470 IFFIIVIMIRAVKANK 487

RESULT 11
C86567
ADP/ATP translocase [imported] - Chlamydia pneumoniae (strain J138)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: C86567
R:Shitai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349; PMID:10871362
A:Accession: C86567
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-540 <STO>
A:Cross-references: GB:BA000008; NID:98978986; PIDN:BA98821.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: act 2
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 39.0%; Score 1025.5; DB 2; Length 540;
Best Local Similarity 39.4%; Pred. No. 6.9e-71;
Matches 214; Conservative 107; Mismatches 185; Indels 37; Gaps 9;

Qy 1 MTKEEPEFGKLRSPFPIHTEHLEKVLPMFPCITFNYTVLRDTOTLIIGARGSGA 60
Db 1 MOSSVKKFSLRAVLCPIYKSEFSKFPPLAFVGVNCLKNMDELIVIGSDAGA 60

Qy 61 EAIPIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPLIFPALTFTVYIPRLDVH 120
Db 61 EVIPIKFWGVPCAVITVTVYGMVLSGRYPADTVFYCFMAAFLGFFFAVITVPGVSL 120

Qy 121 HPTFADRLQALIPGILGLVAIILRMWTFPAFVLAELMGSVMLSLMFGFANEIKIHE 180
Db 121 HNLSTADLQELLPQGLRGFIVMWKYSIYVWSELSVSLMPLWGLANQITITTE 180

Qy 181 AKRFVALFGIGANISILASGRAIYWASKLR--ASVEGDVPM--GISRLMLAMITVSGIVL 238
Db 181 AGRFVALINTGLNLSICAGISYWMKQTFVAASPADSHSVNLNLTMLIT--CSGLIM 239

Qy 239 MASTWINKN--LTPRFNYPEEMOKGKKA-----KPKNNKDSFLYLDSPYI 287
Db 240 I-----MLYRRIHHLITDTSIPSRRLVAEEGAATANKKKKPKAKANLFLHIQRYL 295

Qy 288 LLTTLVLAAGICINLIEVTWSQKLQYPMNDYSEFNGNFSFMTGVVSVLIMFVGGN 347
Db 286 LGLATIVSLVHLHLEFVWKDOVSQIYSSHVEFNGMSAITLIGVSVLAVALLTGQ 355

Qy 348 VIRKGMVLTGALVTPVWVLLTGVIFPALVI--FRNOASGLVAMFGTTPMLAVVGAIONI 406
Db 356 CIRKGMVLTGALVTPVWVLSGLLFFGTIFPAKDISIFGVLGWTPLALAMGMQNV 415

Qy 407 LSKSTKALFSTYEMAYIPLDQOKVKGKAAIDVVAARFGSGGALLIQGLLVI CGSIG 466
Db 416 LSRGKFFPQOTKEMAFIPSPEDKNGKAAIDGVVSRIRKSGSLIYQGLLVI FSSVA 475

Qy 467 AMTYLVALLFFIATMVSATKANKLFLAQA-----LK-----EDEVQOE 508
Db 476 ASLVIALVLLIIMVWIAVAVYIGKEYSRADAVALTKQPKPSSSIVREADESVQOE 535

Qy 509 DSA 511
Db 536 EMA 538

RESULT 12
B72056
ADP, ATP carrier protein CP0133 [imported] - Chlamydia pneumoniae (strains CWL029 an

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: B72056; G81510
R:Katman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: B72056
A:Molecule type: DNA
A:Residues: 1-540 <ARN>
A:Cross-references: GB:AE001646; GB:AE001363; NID:g4376910; PIDN:AD18753.1; PID:g437692
A:Experimental source: Strain CML029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Moyn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: G81610
A:Molecule type: DNA
A:Residues: 1-540 <REA>
A:Cross-references: GB:AE002174; GB:AE002161; NID:g7189059; PIDN:AA838016.1; PID:g718906
A:Experimental source: Strain AR39, HL cells
C:Genetics:
A:Gene: act 2; CP0133
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 39.0%; Score 1025.5; DB 2; Length 540;
Best Local Similarity 39.4%; Pred. No. 6.9e-71;
Matches 214; Conservative 107; Mismatches 185; Indels 37; Gaps 9;

Qy 1 MTKEERKFGKLRSLPFIHTEHLKVLPMFLMFCITFENYTVLRDTKDTLIVAGPSCGA 60
1 MOSSEVKPFSRLKRLCTIKYSEPSKVPFLPLAFVFNCLKMKDTLIVIGSDAGA 60
Db 1 EALPIKFWLVPCALIFMLIYAKLSNLSKQALFYAVGTPELLIFALFPVIYPLRDVL 120
61 EVIDFLKVGIVPGAVITWVYGMVLSRYPDYFCMAFLGFFLFAVITIPVGSGL 120
Qy 121 HPTFADRLQALIPGLIGVAILRNWTFAPFYLAELMSVMSLMEWGPANEITKHE 180
121 HNSLADQLQELPQGLRGIWVRVWYSIYVWSELMSVLSMLFWGLANOITITTE 180
Db 181 AKRFPALFGIGANISLASGRAIYVASKLR-ASVSEGVDP-GISRLMTMTVSGVL 238
181 AGRYALNTLNLSISICAGEISTWKGKOTVAISFACDSWHSVMLNTMLIT-CSGLIM 239
Qy 239 MASYWINKNV-LTDPFYNPEEMQKKGKGA-----KPKNMKDSFLYIDRSPYI 287
240 I-----MIVRRIMHLLTIDISIPPSRVLAEGAAATANLKEKKKPKAKANLFLHLIQSRYL 295
Db 288 LLLTLVLVAYGICINLIVTWKSQLQOYPMNDYSEPMGNFSFMTGVSVLIMLVGNGN 347
288 LLLTLVLVAYGICINLIVTWKSQLQOYPMNDYSEPMGNFSFMTGVSVLIMLVGNGN 347
Qy 296 LGLAIIIVLSYMLVHLFEVWKDQVSQIYSHVEENGYSRITTLIGVSVLAVALTLTGQ 355
296 LGLAIIIVLSYMLVHLFEVWKDQVSQIYSHVEENGYSRITTLIGVSVLAVALTLTGQ 355
Db 348 VIRKFGMTGLVTPVWVLTGIVFPAVI-FRNOASGLVMPFGTTPMLLVVGAIONI 406
348 VIRKFGMTGLVTPVWVLTGIVFPAVI-FRNOASGLVMPFGTTPMLLVVGAIONI 406
Qy 356 CIRKMGMTVGLVTPVWVLTGIVFPAVI-FRNOASGLVMPFGTTPMLLVVGAIONI 415
356 CIRKMGMTVGLVTPVWVLTGIVFPAVI-FRNOASGLVMPFGTTPMLLVVGAIONI 415
Db 407 LSKSTKVALPDSTKEMAYIPLDOEQKVKGAIDVVAARFGSGGALLIOGGLIVICSGIG 466
407 LSKSTKVALPDSTKEMAYIPLDOEQKVKGAIDVVAARFGSGGALLIOGGLIVICSGIG 466
Qy 416 LSRGTGKFFPDQTEMAFIPLSPEDKXNGKALIGVSRICKSGSGLIQGLLVIFSSVA 475
416 LSRGTGKFFPDQTEMAFIPLSPEDKXNGKALIGVSRICKSGSGLIQGLLVIFSSVA 475
Db 467 AMPTVYLAIVLFIATIMLVSAIKNLKFLAOSA-----LK-----EDEVAOE 508
467 AMPTVYLAIVLFIATIMLVSAIKNLKFLAOSA-----LK-----EDEVAOE 508
Qy 476 ASLVNVAIVLVLLIMVWVAIVAYIGKEYYSRAADAVATLKOPKPSSSIVREAGSEVQE 535
476 ASLVNVAIVLVLLIMVWVAIVAYIGKEYYSRAADAVATLKOPKPSSSIVREAGSEVQE 535
Db 509 DSA 511
509 DSA 511
Qy 536 EMA 538
536 EMA 538

RESULT 13
B97790

ADP/ATP carrier protein [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: B97790
R:Ogata, H.; Audic, S.; Renezo-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557693
A:Accession: B97790
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-501 <KRR>
A:Cross-references: GB:AB006914; PIDN:ALU03260.1; PID:g15619815; GSPDB:GN00173
C:Genetics:
A:Gene: tlc3
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 37.1%; Score 976.5; DB 2; Length 501;
Best Local Similarity 40.0%; Pred. No. 3.6e-67;
Matches 199; Conservative 106; Mismatches 174; Indels 19; Gaps 8;

Qy 9 FGKLRSLPFIHTEHLKVLPMFLMFCITFENYTVLRDTKDTLIVAGPSCGAIPPIKF 68
8 FEKVKELMMPDERKELKFLIPMALMLCIFNFGALRSIKDSLIV--PSMGAEIISFELK 65
Db 69 WLVVPCALIFMLIYAKLSNLSKQALFYAVGTPELLIFALFPVIYPLRDVLHPT-EPAD 127
69 WLVVPCALIFMLIYAKLSNLSKQALFYAVGTPELLIFALFPVIYPLRDVLHPT-EPAD 127
Qy 128 RLOALIPGLIGVAILRNWTFAPFYLAELMSVMSLMEWGPANEITKHEARFVAL 187
128 RLOALIPGLIGVAILRNWTFAPFYLAELMSVMSLMEWGPANEITKHEARFVAL 187
Db 126 KLIASV-PNKKWFIKISSQSYALMTIFALIMSVAIVNLMFQPAHNIPTSKKRRYPV 184
126 KLIASV-PNKKWFIKISSQSYALMTIFALIMSVAIVNLMFQPAHNIPTSKKRRYPV 184
Qy 188 FGIGANISLASGRAIYVASKLRASVSEVDP-----WGISRLMTMTVSGVL 239
188 FGIGANISLASGRAIYVASKLRASVSEVDP-----WGISRLMTMTVSGVL 239
Db 185 LGMVGNIGLIIAGSVLIVFSSGQVIDSELDPDSFNSAGNAIMLQIMSIIVAGIIM 244
185 LGMVGNIGLIIAGSVLIVFSSGQVIDSELDPDSFNSAGNAIMLQIMSIIVAGIIM 244
Qy 240 ASYWINKNVLTDPFRFNPPEMOKKGKGAAPKAMKDSFLYIDRSPITLLTLVAYGI 299
240 ASYWINKNVLTDPFRFNPPEMOKKGKGAAPKAMKDSFLYIDRSPITLLTLVAYGI 299
Db 245 LLEPIIRFLITDS-INVLDAKKVTAQMTKLSVIESIKLVHSHKYGIRIALLITCYGL 302
245 LLEPIIRFLITDS-INVLDAKKVTAQMTKLSVIESIKLVHSHKYGIRIALLITCYGL 302
Qy 300 CINLIEVWKSQQLQOYPMNDYSEPMGNFSFMTGVSVLIMLVGNGVIRKFGMLTGL 359
300 CINLIEVWKSQQLQOYPMNDYSEPMGNFSFMTGVSVLIMLVGNGVIRKFGMLTGL 359
Db 303 LNVIEGPMKAKIKELHPNITIDYVNGFRNIMWGISCVFPM-IGSNIIRRLGMLSAL 361
303 LNVIEGPMKAKIKELHPNITIDYVNGFRNIMWGISCVFPM-IGSNIIRRLGMLSAL 361
Qy 360 VTPVWVLTGIVFPAVI-FRNOASGLVMPFGTTPMLLVVGAIONIISKSTKVALPDST 419
360 VTPVWVLTGIVFPAVI-FRNOASGLVMPFGTTPMLLVVGAIONIISKSTKVALPDST 419
Db 362 LTPIMLSITGLMFPFIIFIEICEGCFDFTNL-LVAALIVGAIONITSKSKSLPDS 419
362 LTPIMLSITGLMFPFIIFIEICEGCFDFTNL-LVAALIVGAIONITSKSKSLPDS 419
Qy 420 KEMAYIPLSDLRKGAIVAVIGTRKSGIAGFIOSLFIIPATFDSIIITLITFI 479
420 KEMAYIPLSDLRKGAIVAVIGTRKSGIAGFIOSLFIIPATFDSIIITLITFI 479
Db 420 KEMAYIPLSDLRKGAIVAVIGTRKSGIAGFIOSLFIIPATFDSIIITLITFI 479
420 KEMAYIPLSDLRKGAIVAVIGTRKSGIAGFIOSLFIIPATFDSIIITLITFI 479
Qy 478 FIATIMLVSAIKNLKFL 495
478 FIATIMLVSAIKNLKFL 495
Db 480 VMSIMVIMVINKLKEV 497
480 VMSIMVIMVINKLKEV 497

RESULT 14
B71707
ADP/ATP carrier protein (tlc3) RP477 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: B71707
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichterz-Ponten, T.; Alemark, U
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:9823893
A:Accession: B71707
A:Molecule type: DNA
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Residues: 1-501 <AND>
A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA14932.1; PID:g386103

A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: tlc3; RP477
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 36.9%; Score 969.5; DB 2; Length 501;
Best Local Similarity 39.6%; Pred. No. 1.2e-66;
Matches 197; Conservative 107; Mismatches 176; Indels 17; Gaps 7;

```
OY 9 FGLKRSFLMPPIHTELKLVLPWLMFPCITFNTYVLRDTKDTLIVGAGSGAEALPIK 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 FEKVEKIIMPIERKELKELFIPALMMMLCILFNFGLRSIKDSLIV--PSMGAEIISFKL 65

OY 69 WLWVCAIIFMLIVAKLSNLSKOALFYAVGTFPIIFALPFTVYLPDRDVLHPTFADR 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 WLWVSCVITFITYKLKSNKLNFEIYFISVGTFLFELFPLFYIITYPQDIYHPDANIN 125

OY 129 LQALIPGLGLVALIRMTFAFYVLAELMGVNLMLFMGFANEITRIHEAKRFVALF 188
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 NLIASYPMUKMFIKIGSKMSYALMTIFSELMSAVINIMFQFANHFDLAKAKRFYVL 185

OY 189 GIGANISLLASGRAIVMSKLRASVSEGV--DPM-----GISRLMAAMTIVSGLVMA 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 GMVGNIGILIGSVLVPFSSGGYIIDSELTDSYSSNNNSIMLPITISITVTAGIIMF 245

OY 241 SYMWINKVLDPRFYNEEMOKGKAKPKMMKDSFLYDRSPYILLTLIVAGYIC 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 246 LFRINKETLRNS--INVLDVKAQAATKTKALIESIKLIHSKTYIGRIALLIICYEL 303

OY 301 INLIEVTKSOLKLOYPNNNDYSEFMGNFSFMTGVSVLIMLVGNAVIRKFGMLTGLV 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 304 INIVGPKAKIKELHPPTVDYVFMGMFNIMMGISCTPMI--ISNLIRLGMILISALL 362

OY 361 TPVWVLTGLIYFPALVIRNQAAGLVAMFGTTPMLAVVGAIONILSKSTKVALFDTK 420
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 363 TPIMLSITGFMEFPIIFIEIEIGTCFGDENV--LVVAIIVGAIQNIILSKSKYSLFDTK 420

OY 421 EMAYIPLEDEKMAFPLEDEDEKNGKKAIDVVAARFGSGGALIQGGLVLT--CGSIGAMPYLAIVLTF 478
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 421 EMAYIPLEDEKMAFPLEDEDEKNGKKAIDVVAARFGSGGALIQGGLVLT--CGSIGAMPYLAIVLTF 480

OY 479 IIAIWLVSATKLNKLF 495
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 481 MNMLMIMNIITLNKEYI 497
```

RESULT 15

probable adp/atp translocase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: E71503
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis reference: A71570; MUID:99000809; PMID:9784136
A:Reference number: A71570; MUID:99000809; PMID:9784136
A:Accession: E71503
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-540 <ARN>
A:Cross-references: GB:AE001323; GB:AE002173; NID:g7328931; PIDN:AA068096.1; PID:g732893
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: CT495
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 36.4%; Score 956.5; DB 2; Length 540;
Best Local Similarity 37.2%; Pred. No. 1.3e-65;
Matches 197; Conservative 116; Mismatches 193; Indels 23; Gaps 7;

```
OY 4 TEKPFGLKRSFLMPPIHTELKLVLPWLMFPCITFNTYVLRDTKDTLIVGAPSGAEAI 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 SEVKSFSKRGYFPPIYRSFSPKFIPLFPLAFVGVNVALKTKTDSLIVLGSRAAEVI 62
```

```
OY 64 PFIKFWLVPCAIIFMLIVAKLSNLSKOALFYAVGTFPIIFALPFTVYLPDRDVLHPT 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 PFLKVMGIVGAVLVTIMYGMSSRRYSRGVIFSLVGFLGFPALFATVIYIPIDALHLN 122
```

```
OY 124 EFADRLQALIPGLGLVALIRMTFAFYVLAELMGVNLMLFMGFANEITRIHEAKR 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 KLAQKQSIIPGGRGVVMVQWYSLYVMSLMSIVSLTFMVGVAHHITSVREAGR 182
```

```
OY 184 FYALFGIGANISLLASGRAIVMSKLRASVSEGVDM--GISRLMAAMTIVSGLVMA 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 FYALINIGLMSVFAEVSILMGRNVIAFPNAVDMHEVLNITLLIYLAGGVIL--- 239
```

```
OY 242 YMWINKVLDPRFYNE-----EMOKGKAKPKMMKDSFLYDRSPYILLTL 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 240 YLVKDLRLDETSMLKEGLAEBSVQKKEKRSQAK--AKSLFALLRSRLGLAV 297
```

```
OY 293 LVIAVGCINLIEVTKSOLKLOYPNNNDYSEFMGNFSFMTGVSVLIMLVGNAVIRKF 352
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 298 VVLSYNLVILFEVVMKDQCRVYASRVERFSVMSRITTLTGIVSALTGIFPAAGOTIRW 357
```

```
OY 353 GMLTGAIVTPVWVLTGIVFF--ALVIRNQAAGLVAMFGTTPMLAVVGAIONILSKST 411
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 358 GMTIGALVPLVMTLITGALFFGAIYAVGDMIFGILGISPLVTAMLGQVQVFSRAI 417
```

```
OY 412 KYALFDSTKEMAYIPLEDEKNGKKAIDVVAARFGSGGALIQGGLVTCGSI GATPY 471
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 KFTTFDDTKEMAFPLEDEDEKNGKKAIDVVAARFGSGGALIQGGLVTCGSI GATPY 477
```

```
OY 472 LAVILFIIAIVLSATKLNKLF 514
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 478 ITVILALISGWFVIMLGREVYAKTEALFRVNVSEEDVLQERREASS 526
```

RESULT 16

ADP, ATP carrier protein TC0782 (imported) - Chlamydia muridarum (strain N19g)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C:Accession: F81665
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, J. C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: F81665
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-543 <RET>
A:Cross-references: GB:AE002346; GB:AE002160; NID:g7190805; PIDN:AAF39585.1; PID:g719080
A:Experimental source: strain N19g (MoPn)
C:Genetics:
A:Gene: TC0782
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 36.2%; Score 953; DB 2; Length 543;
Best Local Similarity 37.5%; Pred. No. 2.5e-65;
Matches 199; Conservative 111; Mismatches 197; Indels 24; Gaps 7;

```
OY 4 TEKPFGLKRSFLMPPIHTELKLVLPWLMFPCITFNTYVLRDTKDTLIVGAPSGAEAI 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 SEVTFKFRFRYFPPIHKSSEFPKFIPLLLAFVGFVSLTKTKDSLIVLAGSRAAEVI 62
```

```
OY 64 PFIKFWLVPCAIIFMLIVAKLSNLSKOALFYAVGTFPIIFALPFTVYLPDRDVLHPT 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 PFLKVMGIVGAVIITIMYGMSCRSRGVFCALVGGSFPAFPALFCVITYPMDALHLN 122
```

```
OY 124 EFADRLQALIPGLGLVALIRMTFAFYVLAELMGVNLMLFMGFANEITRIHEAKR 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 GLAAKQDTILPRAGRGVVMVQWYSLYVMSLMSIVSLTFMGIANHITSVREAGR 182
```

```
OY 184 FYALFGIGANISLLASGRAIVMSK--LRASVSEGVDM--GISRLMAAMTIVSGLVMA 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 184 FYALFGIGANISLLASGRAIVMSK--LRASVSEGVDM--GISRLMAAMTIVSGLVMA 241
```


Db 183 FYALINVLGIVAGSIVAGEISLWIKGKTLIPSSMAVDAMHGVLLNTLLIIVAGGHL--- 239

Qy 242 YWMINKNTLDP-----RFYNPEEOKGKKAKPKMMKDSLYLDRSPYLLTL 292

Db 240 YLYRKLIDHTEBAPVLDGLVSEMSVAQKQEK--RPRAKAKSLSLVYFRSRYIMGLAV 297

Qy 293 IIVAYGICINLIEVTWKSQKLQYPMNDYSEFMGNFSFMTGVSVLYLMPVGNVIRKF 352

Db 298 VVLAINLAIHLEAVWKEVQCIYSRVSFNSYMSITFTGIVSALAGVFAAGSIRRW 357

Qy 353 GMLTALVTPVAVLLTGIVF--ALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKST 411

Db 358 GWTVALITPLTMLITGLFPGAIYAVKGDAMILGGLFSPFLVTLAMLGIVQNVESRAI 417

Qy 412 KVALDSTKEMAYIPLDQOKYKGRALIVVAARPKSGGALIOGLIVCGSIGAMPY 471

Db 418 KTYPDQTEKAFIPLEDEKQYGRALIDGIVSRVKSQSGSLVYQALLIFSSVADCMNA 477

Qy 472 LAVILFIIAIVLVSATKLKLFQAQS-----LKQEQVNAQEDSAPASS 515

Db 478 ITIVILLALGMIWVAVMLGKEYSVRTALGKRAAEPSLDEDESRVSS 528

RESULT 17

B97842
ADP carrier protein (imported) - Rickettsia conorii (strain Malish 7)

C:Species: Rickettsia conorii

C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001

C:Accession: B97842

R:Ogata, H.; Audic, S.; Reneato-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro Science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.

A:Reference number: A97700; PMID:21442074; PMID:11557893

A:Accession: B97842

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-499 <KUR>

A:Cross-references: GB:AE006914; PIDN:AA03676.1; PID:g15620264; GSPDB:GN00173

A:Gene: C1G5

C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 34.5%; Score 907.5; DB 2; Length 499;

Best Local Similarity 37.4%; Pred. No. 6.9e-62;

Matches 190; Conservative 107; Mismatches 176; Indels 35; Gaps 13;

Qy 1 MTKTEKDF-GKLRSFLMPTITHELKLVLPMLFPCITFNTVTLRDTKTLIVAGPSSG 59

Db 1 MLSTSSRSFQNFRAAFMVFVHVELGKFIPISTLMFCILFQNVLRILKDSILISE--IS 58

Qy 60 AEAIFPIKFWLVPCALIFMLIYAKLSNLSKQALFYAVGTEPFLIPALFPIYVPLNDV 119

Db 59 AEAIAFAKYCVTPAALFVLIYAKMINVLTPEKIFYYLSAFFISFVLFTEVIYINHI 118

Qy 120 --LHPTFADRLQALIPGLGLVALIRWTFAPFVLAELMGSVMLSMFGFANEITK 177

Db 119 FTHVFNRLADWERY--PHEKMYISLVGNGYIVYISLAELMPLNFYVLLFQFANELT 176

Qy 178 IHEAKRFYALFGIGANISILASGRAIWMAS-----KLASVSEG--VDPWGISLRLMAM 230

Db 177 TEBAKRFYTLFSLFQNSGLILVGFPMNONSSEDITIKKMSISDSKITLVQSTTIVAI 236

Qy 231 TIVSGVLMASTWYMNKQVLDPRRYNPEEOKGKG--AKPKAMKDSFLYLDSPYIL 288

Db 237 AIIICCLVRA---FISKVFTNPLFY---AKASGRSTSERMGLIKSFYIAKSKYLM 287

Qy 289 LITLIVAYGICINLIEVTWKSQKLQYPMNDYSEFMGNFSFMTGVSVLYLMPVGNV 348

Db 288 LITLISAAFGPAIVLVEAVMKAKIELPTVNTVAFPSLYILMGV--VAIWMITIIIGNNI 346

Qy 349 IIRFGMLGALVTPVAVLLTGIVFPAVIFRQA-----SGLVAMFGTTPMLAVVGAIO 404

Db 347 KRMHMFVAAVISPIYIMVTGILFVLIYFDQIISLFDGAILM--SPLALAVSIGIQ 403

Qy 405 NILSKTYALPDSKEMAYIPLDQOKYKGRALIDVVAARPKSGGALIOGL--LYIC 462

Db 404 NILAKGTYSTWDSIREMLYIPLDEELTKGAADVIAKVGSSSGVOSITFTLVPT 463

Qy 463 GSIGAMTPYLAIVLFIIVLVSATKL 490

Db 464 ATFTLISPLIWWVFVFLAMIVAVRKI 491

RESULT 18

B97783
ADP carrier protein (imported) - Rickettsia conorii (strain Malish 7)

C:Species: Rickettsia conorii

C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001

C:Accession: B97783

R:Ogata, H.; Audic, S.; Reneato-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro Science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.

A:Reference number: A97700; PMID:21442074; PMID:11557893

A:Accession: B97783

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-511 <KUR>

A:Cross-references: GB:AE006914; PIDN:AA03204.1; PID:g15619755; GSPDB:GN00173

A:Gene: C1G4

C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 34.5%; Score 907.5; DB 2; Length 511;

Best Local Similarity 37.9%; Pred. No. 7.1e-62;

Matches 187; Conservative 111; Mismatches 175; Indels 21; Gaps 9;

Qy 9 FGKLSFLMPTITHELKLVLPMLFPCITFNTVTLRDTKTLIVAGPSSGALIPFIKF 68

Db 19 FSKLDIYMPIKREHVSKEFLITLMFCILFIQNLIRALKDSIVTWM--IGAEIISFLKF 76

Qy 69 WLVPACALIPMLIYAKLSNLSKQALFYAVGTEPFLIPALFPIYVPLNDVH--PTEFA 126

Db 77 WGVESAPLMTAIVLKVLRKKAENIFLIISITLFPALFAYVIFPHMELHESPVTQ 136

Qy 127 DRLQALIPGLGLVALIRWTFAPFVLAELMGSVMLSMFGFANEITKHEAKRFA 186

Db 137 NLMASL--PNLKWPIWLSKMSFSFLYIAILMERNVALLFMQFVNNTTVEBSKREYP 194

Qy 187 LFGIGANISILASGRAIWMASKLASVSEGVDPWG---ISRLMAMTIVSGVLMAST 242

Db 195 LFGILSQGIVLAGQFLENLSNINDVYTNKFAQSFTLSIQLITVLILGIAIKTF 254

Qy 243 WINKNVLTDRFPYNPEEOKG--KGAKPKAMKDSFLYLDSPYILTLIVAYGICI 301

Db 255 WLNHKKVL-----DKEMALLRFKAKKSKWTIASEFMLSSRIIRIATLLCYGLAI 308

Qy 302 NLIEVTWKSQKLQYPMNDYSEFMGNFSFMTGVSVLYLMPVGNVIRKFGMLGALVT 361

Db 309 NLVSGPMFAATKYKPTETEAALIGSLYSTIGFTIIFV--LGSNIYRKLGMPTAAVIT 367

Qy 362 PVMVLLTGIVFPAVIFRQASGLVAMF--GTTPLMAVAVGAIONILSKSTYALPDSK 420

Db 368 PLIVITGILFPAVNNFRPAGLIIANLILDPALIAITIGAIQVLSKSKYTLFDSK 427

Qy 421 EMAYIPLDQOKYKGRALIDVVAARPKSGGALIOGLIVCGSIGAMTPYLAIVLFI 480

Db 428 EMAYVPLDPEIKIGKKAADVIGTKGSGSAPLOSIVFIILPSSAYOSISTCIMIIFI 487

Qy 481 A--IWLVSATKLAK 492

Db 488 TGLTLMATKALNK 501

RESULT 19

F71653
ADP carrier protein (C1G4) RP500 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: F71653
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Almark, U.
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71650; MUID:99039499; PMID:982893
A:Accession: F71653
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Gene: tlc4; RP500
C:Superfamily: rickettsial-type ATP/ADP translocase .

Query Match 33.5%; Score 880.5; DB 2; Length 512;
Best Local Similarity 36.9%; Pred. No. 8.3e-60;
Matches 184; Conservative 108; Mismatches 177; Indels 29; Gaps 10;

QY 9 FGKRSFLMPPIHTEHKLVKLPMLFPCITFNNYTVLRDXTDLIVGAGSGAIPAIRK 68
DB 19 FSKLTDYIWPPIKRBHISKFLFTLLMFCILFIQNLRLKDSIYTM--IGAEITSLKF 76
QY 69 WLVPFCALIFMLIYAKSNILSKQALFYAVGTPFLIFALFPYIYPLRDYLH--PFEFA 126
DB 77 WGVPSAPLLIVIVYKLVNRMKAENIFYLIIISFLTFPALFAYVIFPHEMLHRPVT-V 135
QY 127 DRLOALDPGLGLGVALIARNWTFPAFYVLAELMGSMVLMLFMGPANBITKHEAKDFA 186
DB 136 HNLRLSL-PNLKWFLLLSKMSFSLFTIILMELPNVVALFLFQPNNTTVESSKRFYP 194
QY 187 LFGIGANISLISAGRAIYVASKLRASVEGVDPMG---ISLRLLMAMTIVSGVLNMA 242
DB 195 LFGILSQGIYLAHGFLENLSINIVYVTKRKALQSPFTTISIQIILTVLILGIVSKTF 254
QY 243 WVINQVLTDP----RYNPEBNQKGGKAKPKKMKDSTLYIDRSPIYILLTLVIAY 297
DB 255 WLNHRKVLDDKGMALLRP-----KTKNKSITIAKSPQMTLSRHRIRLITLLICY 304
QY 298 GICINLIEVTWKSQOLKQYPMNDYSEPMGNSFWTGVSVLMLFVGANVIRKGMLTG 357
DB 305 GIALLNVEGPKMAATKTKYKTPTEYAAFTGISLSTGVFTTFVYL-IGSNIVRMGFIS 363
QY 358 ALVTPVWMLLTGIVFALVIFRNQASGLVAAF-GTTPPLMLAVVGAIONILSKSTKALF 416
DB 364 AVITPSIVFTIGILFPAANNFEGFAGLIIAFIILDPALVALITIGALQNVLSKSKXTLF 423
QY 417 DSTKEMAYIPLDQOKVKGKAAIDVVAARFSGAGALIQGLLVIGSIGAMPEYLAAIL 476
DB 424 DSTKEMAVPLPEPEIKTISGKAADVIGTKLGSAGSAFLQSLIFILPSASYSISICMI 483
QY 477 LFIIL-ATWLVSATKLNK 492
DB 484 IFILTCVTWATKELNK 501

RESULT 20
G71653
ADP/ATP carrier protein (tlc5) RP739 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: G71653
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Almark, U.
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71650; MUID:99039499; PMID:982893
A:Accession: G71653
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-500 <AND>
A:Cross-references: GB:AJ235273; GB:AJ235269; NID:g3861033; PIDN:CAA15167.1; PID:g386124

A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: tlc5; RP739
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 32.9%; Score 865; DB 2; Length 500;
Best Local Similarity 35.7%; Pred. No. 1,2e-58;
Matches 176; Conservative 109; Mismatches 182; Indels 26; Gaps 10;

QY 11 KLSFLPPIHTEHKVLPWFLEFCITENVTVLRDTKDTLIVAGPSSGAEALPFIKFWL 70
DB 13 KFRFAFPVHNVEYLGKPIKISALMFICITLQNILIRIKIDSLISE--ISAEINGFAKYC 70
QY 71 VVPCAIIFEMIIVAKLSNLSKQALFVAGNPFLIFALPFTVIVPLSDV--LHPTEADR 128
DB 71 VTPAALFVLIIVAMINHLTFEKFYYLSAFETSCFLFPAFVIVPNIHVHVDITLSDW 130
QY 129 LQALIPGLGLVALILNNMTFAAFVYLAELMGSYMLSLMEFGPANEITKHEAKRFPALF 188
DB 131 MNKTL--EHEKRYISLVGNMGVIVYSLAEILMPNIFVLLRMOFTNELTTEAKRFTLTF 188
QY 189 GIGANISLLASGRAIIVASKLRASVSEGVDPWGISRLMLAMTIVSGLV----LMA5Y 243
DB 189 SLFNGSSLIIVGLFMNMLSSBDITIKKFIISDSKITLVGVSTPIIAIVAILICCLAVRF- 247
QY 244 WINKVULDPREFVPEEMQKKGAKRNNMKCSFLYLDSPYITLLTLIVAYGICINL 303
DB 248 -ISKRYITNPLEYH--KTKSSRSTAGRMGLIKSFKIVKSKYLMILLICSAAFGEFAINL 303
QY 304 IEVVMKQOLKQYNNNDYSEFMGNFSPMTGVVSVLMLFVGVGNVIRKFGMLTGAUTVPV 363
DB 304 VEAVMKAKIKELVPTVTVTAFFNSLYILMTG-VAILMTIIGNNNVRMHHVFAAVISPV 362
QY 364 WVLITGVIFPALVIFRNAQ---SGLVAMFGTTPMLAVVGAIONLSKSTYALFDST 419
DB 363 IIMVTGVLFGELIVFDQIILSTLFDGAILM--SPLALAVISGGIONILAKGTYSIWDTS 419
QY 420 KEMAYIFLDOQKKGKGAALIDVVAARFGSGGALIQGL--LVICSGIGMTFYLAIVILL 477
DB 420 REMYIYFLDELKTKRGAADVISAKEYKSSSGLVOSIIFTLVPMATFTSISPLMVFVT 479
QY 478 FIIMVLEVSATKL 490
DB 480 FVCFAMVYAVRKI 492

RESULT 21
B97765
ADP,ATP carrier protein (imported) - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: B97765
R:Ogata, H.; Audin, S.; Renseto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro-
science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; M01D:2142074; PMID:11557893
A:Accession: B97765
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-507 <XLR>
C:Genetics:
A:Gene: tlc2
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 31.4%; Score 827; DB 2; Length 507;
Best Local Similarity 35.8%; Pred. No. 1e-55;
Matches 179; Conservative 106; Mismatches 183; Indels 32; Gaps 9;

QY 11 KLSFLPPIHTEHKVLPWFLEFCITENVTVLRDTKDTLIVAGPSSGAEALPFIKFWL 70
DB 11 KFRFAFPVHNVEYLGKPIKISALMFICITLQNILIRIKIDSLISE--ISAEINGFAKYC 70
QY 189 GIGANISLLASGRAIIVASKLRASVSEGVDPWGISRLMLAMTIVSGLV----LMA5Y 243
DB 189 SLFNGSSLIIVGLFMNMLSSBDITIKKFIISDSKITLVGVSTPIIAIVAILICCLAVRF- 247
QY 244 WINKVULDPREFVPEEMQKKGAKRNNMKCSFLYLDSPYITLLTLIVAYGICINL 303
DB 248 -ISKRYITNPLEYH--KTKSSRSTAGRMGLIKSFKIVKSKYLMILLICSAAFGEFAINL 303
QY 304 IEVVMKQOLKQYNNNDYSEFMGNFSPMTGVVSVLMLFVGVGNVIRKFGMLTGAUTVPV 363
DB 304 VEAVMKAKIKELVPTVTVTAFFNSLYILMTG-VAILMTIIGNNNVRMHHVFAAVISPV 362
QY 364 WVLITGVIFPALVIFRNAQ---SGLVAMFGTTPMLAVVGAIONLSKSTYALFDST 419
DB 363 IIMVTGVLFGELIVFDQIILSTLFDGAILM--SPLALAVISGGIONILAKGTYSIWDTS 419
QY 420 KEMAYIFLDOQKKGKGAALIDVVAARFGSGGALIQGL--LVICSGIGMTFYLAIVILL 477
DB 420 REMYIYFLDELKTKRGAADVISAKEYKSSSGLVOSIIFTLVPMATFTSISPLMVFVT 479
QY 478 FIIMVLEVSATKL 490
DB 480 FVCFAMVYAVRKI 492

QY 71 VPCALITFMLIYAKLSNLSKQALFYAVGTFELIFPVLIVLRDLVLPTEFADRLQ 130
 Db 76 EMPGILFVIFYSLCNIMTEQVFRITITGTFELFFALFGLFPEHFEHDELBKH 135
 QY 131 AILPGLGLVAILRNNTPFAFYVLAELMGSVMSLMFWGFANEITKHEAKREYALFGI 190
 Db 136 ITVLPHLKWFLIINGQMSLVIFYINGELMPVIFTLVWQLANKITKVEAPREYSPFTL 195
 QY 191 GANISILASGRAIYVASK-----LRASVSEGVDPWGISIRLLMAMTIVSGVLMASTYMIN 246
 Db 196 FQGTNLLISGVIIITFYPAKSEHFLPLPESHNDINEILKSPITVILISGLICLALHKLID 255
 QY 247 KNLVLTDRFYNPEEMQKGGKAKP-----KAMNKSFLYDRSPYIILLTLVIAVGICI 301
 Db 256 KSVV-----EADNITFKKQORDILKLSVDSKAVILTSRYLGFICLIVMSYSMSI 306
 QY 302 NLIEVTWKSQKLOQYPMNDYSEFMGNFSFWTGVSVLIMLVGNAVIRKFWMLTGLAVT 361
 Db 307 SLIEGLMMSKVQKQLYPATKDFIAYHGKVFMTGILT- LVSAFLGSSILRIIGWFMGAIIIT 365
 QY 362 PWWALLTGIVFPALVIRNQAAGLVAMFG-TTPEMLAVVGAIONILSKSKYALFEDSTK 420
 Db 366 PIMMFGAGVWFSEFTPEHNLGNITVLTGYSAPLVIVITGGLMHWLSKVYSLFPATK 425
 QY 421 EMAYIPLDPOKQVKGKAIIDVVAARFGKSGALIQGGLVTCGSI--GAMTPYLAIVLIF 478
 Db 426 EMYIYIPDSEKTKGKAADVMAKIGKISGAILIQ----FISFIFPNAIINDIAGLIMF 481
 QY 479 ----IAIWLVSATKLNKL 494
 Db 482 SFIVCLMLYGVKVLKSY 501

RESULT 22

B71695
 A:ATP carrier protein (tlc2) RP377 - Rickettsia prowazekii
 C:Species: Rickettsia prowazekii
 C>Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
 C:Accession: B71695
 R:Anderson, S.G.B.; Zomorodipour, A.; Anderson, J.O.; Sichevitz-Ponten, T.; Almark, U.
 Nature 396, 133-140, 1998
 A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
 A:Reference number: A71630; MUID:99039499; PMID:9823893
 A:Accession: B71695
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-507 <AND>
 A:Cross-references: GB:AJ235271; GB:AJ235269; NID:G3868717; PIDN:CAA14836.1; PID:G386093
 A:Experimental source: strain Madrid E
 C:Genetics:
 A:Gene: tlc2; RP377
 C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 31.3%; Score 824; DB 2; Length 507;
 Best Local Similarity 34.2%; Pred. No. 1.7e-55;
 Matches 175; Conservative 114; Mismatches 170; Indels 52; Gaps 10;
 QY 11 KLRSLPMTHTHEKLVPMFLMFCITFNVTVLADTKDILLVAGPSGAAIDPIKFWL 70
 Db 18 KFRHIVPDRSYELTKFIPMTLMLFILLNONSIDSFVLLISS--EVLISFKLWG 75
 QY 71 VVPCALITFMLIYAKLSNLSKQALFYAVGTFELIFPVLIVLRDLVLPTEFADRLQ 130
 Db 76 EMPGILFVIFYSLCNIMTEQVFRITITGTFELFFALFGLFPEHFEHDELBKH 135
 QY 131 AILPGLGLVAILRNNTPFAFYVLAELMGSVMSLMFWGFANEITKHEAKREYALFGI 190
 Db 136 ITVLPHLKWFLIINGQMSLVIFYINGELMPVIFTLVWQLANKITKVEAPREYSPFTL 195
 QY 191 GANISILASGRAIYVASK-----LRASVSEGVDPWGISIRLLMAMTIVSGVLMASTYMIN 246
 Db 196 FQGTNLLISGVIIITFYPAKSEHFLPLPESHNDINEILKSPITVILISGLICLALHKLID 255

QY 244 WINKNVLTDRFYNPEEMQKGGKAKP-----KAMNKSFLYDRSPYIILLTLVIAVG 298
 Db 253 LIDKSVV-----EADNITFKKQORDILKLSVDSKAVILTSRYLGFICLIVMSYS 303
 QY 299 ICNILEVTWKSQKLOQYPMNDYSEFMGNFSFWTGVSVLIMLVGNAVIRKFWMLTGLA 358
 Db 304 MSINLEGLMMSKVQKQLYPATKDFIAYHGKVFMTGILT- LVSAFLGSSILRIIGWFMGA 362
 QY 359 LVTPWVALLTGIVFPALVIRNQAAGLVAMFG-TTPEMLAVVGAIONILSKSKYALFED 417
 Db 363 IITPIMMFGAGVWFSEFTPEHNLGNITVLTGYSAPLVIVITGGLMHWLSKVYSLFP 422
 QY 418 STKEMAYIPLDPOKQVKGKAIIDVVAARFGKSGALIQ-----QGLLVIC 462
 Db 423 ATTEMYIYIPDSEKTKGKAADVMAKIGKISGAILIQFISFIFPNAIINDIAGLIMF 479
 QY 463 GSIGAMTPYLAIVLIFIAIWLVSATKLNKL 493
 Db 480 ----WVTFIVCLMLYGVKVLKSY 504

RESULT 23

H82642
 A:hypothetical protein XFI738 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: H82642
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:2035717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: H82642
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-441 <SIM>
 A:Cross-references: GB:AE003997; GB:AE003849; NID:G9106805; PIDN:AAF84547.1; GSPDB:GN001.
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, H.
 as-Neto, E.; Docena, C.; El-Dodry, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to Genbank, June 2000
 A:Authors: Ferreira, V.C.A.; Perro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm.
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitejina, J.P.; Krieger, J.E.; Kurame, E.E.; Laig.
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir.
 M.; Tsubako, M.H.; Valada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Za
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XFI738

Query Match 8.7%; Score 229.5; DB 2; Length 441;
 Best Local Similarity 23.1%; Pred. No. 5e-10;
 Matches 122; Conservative 84; Mismatches 197; Indels 125; Gaps 27;
 QY 3 KTEKPKGLKRSFLMPTHTHEKLVPMFLMFCITFNVTVLADTKDILLVAGPSGAA 62
 Db 22 RPEAP-----AVLM-----SMLYVALFLAY-----VLRPIREL--GVAG--GVQN 61
 QY 63 IPIKFLVVPVCAIIFMLIYAKLSNLSKQ-----ALFYAVGTFELIFPVLIVR 115
 Db 62 LP-----WLFTG--TLTAMLVASPLFALAVRSIPROFIALAVRFAALVIFAL----- 109
 QY 116 LRVDTHTERADLUQALPGLGLVAILRNNTPFAFYVLAELMGSVMSLMFWGFANEI 175
 Db 110 ----LLH-----FADPQWQV-----WGRAPFIWVSFNLFWVS--VFASFVVDL 148
 QY 176 TKIHEAKREYALFGI GANISILASGRAIYVASKLRASVSEGVDPWGISIRLLMAMTIV- 233

Db 149 FDSEGRKLFGLFAAGAT----AGS---LGSATISGLIHLDREW-----LMAIAVIF 195
 Qy 234 SGLVIMASYMMINXVLTDPFY-----NPEBQKGGKAKPKKMKDSFLYIDRSFYI 287
 Db 196 LELAVIAS-----RRLSRIPAFQHAARDNPDPGLG-----GIPAGWHILRSPLYL 243
 Qy 288 LLTLTLVAVAGCINLIEVTKWSQKLOYPNNNDSEFMGNFSFTGVSVLIMFVCGN 347
 Db 244 GGLAIFILLVSVTSTFLYFQOASIAQASFPPDAATAPAFANIDLVNAVTVLFQLEFVTR 303
 Qy 348 VIRKGMVLGALVTFPMVLTGTIVFPALVIFRNQASGLVAFMGFTPLMAVVGAIQNL 407
 Db 304 MMATVGIATLCVLPVLELL-----GFAAL-ASP-SVAIVIVA--QVA 343
 Qy 408 SKSTKYALFDSTKEMAYIPLDQKVKKAIDVVAARFGSGGALIQOGLIVICSGIGA 467
 Db 344 RRVANFALARPARELFTSSAREDDYKAKNFIDTVVYGGQIISWGYAGLMGL-GLTIA 402
 Qy 468 MTPYIAVLLFTIAIWLVSATKNTLFLAQSAKKEQVADSDAPASS 515
 Db 403 QIPMLAVPL---SAVWLGLSVWLGRTHQAQE-----RQDAATAPS 439

RESULT 24

hypothetical protein CT234 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C/Species: Chlamydia trachomatis
 C/Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 09-Jun-2000
 C/Accession: F71541
 R/Stephen, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
 A/Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
 A/Reference number: A71570; MUID:99000809; PMID:9784136
 A/Accession: F71541
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-928 <ABR>
 A/Cross-references: GB:AE001296; GB:AE001273; NID:93328630; PIDN:AA67826.1; PID:9332864
 A/Experimental source: serotype D, strain UW-3/Cx
 C/Genetics:
 A/Gene: CT234
 C/Superfamily: Chlamydia trachomatis hypothetical protein CT234

Query Match 6.6%; Score 173.5; DB 2; Length 928;
 Best Local Similarity 19.9%; Pred. No. 2.2e-05;
 Matches 106; Conservative 88; Mismatches 195; Indels 143; Gaps 22;

Qy 7 KPFGKLRG--FLMPHTHELKVLPMFLMFCITFNVTVLBDTKOTLIVGAPGS----- 58
 Db 6 KGGSLRLRLRLLSIRGEEKAL-LFLLGLI-----WSVACYSIALGES 51
 Qy 59 -----GAEALP-----IKFWLVVPCAIIFMLIVAKSLNISKOALFYAVGTPPLI----- 104
 Db 52 LFLBIEIGAEKLPFAVLGASFPL---CFISCLILVNLRSKRVSPKALFLSFISCVLICLY 108
 Qy 105 -PFAIFPVIVYPLRDVHLPTERADRLQALIRPGLIGLVAIIRNMTFAAFYVLAELMG-SV 162
 Db 109 LFMHL-----AIH-----KGVSGTPFL-----YRLI--IWGLTI 136
 Qy 163 MLSLMFMGFANETIKHEAKRFVALFGIANISLASGAIYVMAK--LRAVSEGVDPWG 221
 Db 137 LCYANFMGFIQFPIQAKGHFCIF-----NALTFCGDFGAIIVNOIQYLG 184
 Qy 222 ISLRLLAMTIVSGVLVMAASYWINKVLTDR-----FYNPEBQKGGKAKPKKMMK 275
 Db 185 AELILARIYITVIFPLVHTYISSLSKELSEHDHFLDTGYPPSTKQ-----TLKICLK 238
 Qy 276 DSFLYLDSPYILLTLVIAVGICINLIEVTKWSQKLOYPNNM--DYSEFMGNFSFT 333
 Db 239 DKYTFYLVSYFLMQLLVFT-----ERNYKIFPAQCGMETCETLTENFTKYSWI 290
 Qy 334 GVAVVLVIMLFVGNVIRKEGWLGTALVTPVNVLLTGIVFPALVIFRNQASGLVAFMGFTTP 393

Db 291 SLGNMFPALPAYSVRITRFG-----INNILFAPICF-----FSLFCQWS 330
 Qy 394 LMLAVVGAIONISKSTKYALPDSTKEMAYIPLDQKVKKAIDVVAARFGSGGAL 453
 Db 331 IKTSVFATMGMIAREGLAVALDDNNLQLLIYGPKNIRQVRAIE-----SF 379
 Qy 454 IQOGLVICSIGMTPYLAIVILFIATWLVSA-----TLNKLFLAQSAK 501
 Db 380 VEPAGMFCALLCLFPHQYVLCIISAVCILALIRTHYSKAILNLSIE 431

RESULT 25

conserved hypothetical protein TC0505 (imported) - Chlamydia muridarum (strain N:9g)
 C/Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 20-Jun-2000
 C/Accession: B81695
 R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, F.C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A/Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
 A/Reference number: A81500; MUID:20150255; PMID:10684935
 A/Accession: B81695
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-918 <TRT>
 A/Cross-references: GB:AE002319; GB:AE002160; NID:97190541; PIDN:AAF39347.1; PID:9719054
 A/Experimental source: strain N:9g (MoPn)
 C/Genetics:
 A/Gene: TC0505
 C/Superfamily: Chlamydia trachomatis hypothetical protein CT234

Query Match 6.1%; Score 161; DB 2; Length 918;
 Best Local Similarity 19.0%; Pred. No. 0.0002;
 Matches 88; Conservative 80; Mismatches 198; Indels 96; Gaps 14;

Qy 57 GSGEALPFIKPMVVPICAIIFMLIVAK---LSNIISSKOALFYAVGTPPLIFALPFTVI 113
 Db 13 GEEKRALFLTLGLWSVACYSIALGESLFLBIEIGTKLPFAVLGASFPLCF-ISCVL 71
 Qy 114 YPLRDVHLPTERADRLQALIRPGLIGLVAIIRNMTFAAFVLA-----EL 158
 Db 72 YSL-----SRKASPKALFLSFISCVLCNCLVLFPHLAIHRGVSGTPFLVRIIL 121
 Qy 159 WG-SVMLSMFMGFANETIKHEAKRFVALFGIANISLASGAIYVMAK--LRAVSEGV 216
 Db 122 WGLTILCYANFMGFWVDQFNIQAKRHFCIF-----NALTFCGDFGARIYVQ 169
 Qy 217 VDPWGISLRLLAMTIVSGVLVMAASYWINKVLTDR-----FYNPEBQKGGKAKP 270
 Db 170 IQHLGAEILILSFIAVITFPFLVHTYISSLSKELSEHDHFLDTGYPPSAKQAF----- 223
 Qy 271 KMMNKDSFLYLDSPYILLTLVIAVGICINLIEVTKWSQKLOYPNNNDY--SEFMGN 328
 Db 224 KLCKDKYTFYLVSYFLMQLLVFT-----ERNYKIFPAQCGNANKELTENFTK 275
 Qy 329 FSEFTGVASVILMLFVGNVIRKEGWLGTALVTPVNVLLTGIVFPALVIFRNQASGLVAM 388
 Db 276 YSSWISLGNMFPALPAYSVRITRFG-----INNILFAPICF-----FSL 315
 Qy 389 FGTPPLMLAVVGAIONISKSTKYALPDSTKEMAYIPLDQKVKKAIDVVAARFGK 448
 Db 316 FVCKSITSVILNTMGMIAREGLAVALDDNNLQLLIYGPKNIRQVRAIE----- 367
 Qy 449 SGGALIQOGLVICSIGMTPYLAIVILFIATWLVSAATKL 490
 Db 368 ---SFVEPAGMFCALLCLFPHQYVLCIISAVCILALIRTHYSKAILNLSIE 406

RESULT 26

A72096
 ct234 hypothetical protein - Chlamydia pneumoniae (strain CWL029)
 C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000

C:Accession: A72096

R:Kakman, S.; Mitchell, W.; Martche, R.; Lammell, C.; Fan, J.; Olinger, L.; Grimwood, J.;

Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*.

A:Reference number: A72000; MUID:99206606; PMID:10192388

A:Accession: A72096

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-925 (AAR>)

A:Cross-references: GB:AE001614; GB:AE001363; NID:G4376562; PIDN:AD18442.1; PID:G4376566

A:Experimental source: strain CWL029

C:Genetics:

A:Gene: CPn0293

C:Superfamily: *Chlamydia trachomatis* hypothetical protein CT234

Query Match 6.0%; Score 158.5; DB 2; Length 925;

Best Local Similarity 20.6%; Pred. No. 0.00031;

Matches 95; Conservative 94; Mismatches 16; Indels 109; Gaps 26;

DB 59 GAELPFIKFWL---VPECAIFEMLIYAKLSNLSKQALF-YAVGTPPLIFPALFPVIY 114

49 GSAELP-KIYLGSSLLICVLSLILYMLPKKHISATLFLIPVSLISCNFVILISIF 106

DB 115 PLRDLVHTEFERADRLQALLPFGELGVVAILENTPAAYVLAELMGSYMLS-LMFWGFAN 173

107 -----AIDPP-----RSPPLF-FYRIY-IWSLILSYTSFWSGFVD 138

DB 174 EITKHEAKRFYALFGIGANISLASGRAYWASGLRASVSEGVDPWGISRLMAMTV 233

139 QFNLQDGRKHCITN-----AIFLGDIH-GSSIISLVHTIGIQLILFLPALVLT 191

DB 234 SGLVLMASVWYINKV--LTDPR--FYNPEWQKQKQKAKPRMMKQSFYLDSPYIL- 288

192 FPIV-----FVYSKSLKSLSDHDLFID-----TGHPPPLSKALKLCFYDKTYFYL 238

DB 289 -----LTLTLVAVGICINLIEVTWKSQKLQTPMNDYS--EENGNBSFMTGVSVLIML 342

239 CFYFMQLLAIA-----TEFNLIKPEIOPASKEPELVAHIGKSLWISLGNMCPAL 291

DB 343 FVGGVNIKRFQMLTGLALTVPWVLLTGIVGFALVIFRQASGVAMRGTPPLMLAVVGA 402

292 FAYSHIVKRLG-----VNNIILFAPLDFLS-----LFLWTFPTTSLIAVLA 333

DB 403 IONTISKSTKYALPDSTKEMAY--IPLDOEKVKYKKAIDVVAARFGSGALLIQGILV 460

334 M--VVRGVTYALDDNNQLILTYGP-----KIRNQ--IRIVDSFIEPIGLVMS---L 382

DB 461 ICGSIGAMTPYLAIVLFLFIAMVWSATKMLFLAQSALK 501

383 IC--FLSSQOYVFCIIISLIATIIVC--LVRSYAKAILK 418

RESULT 27

EB1573

conserved hypothetical protein CP0465 [imported] - *Chlamydia pneumoniae* (strain AR39)

C:Species: *Chlamydia pneumoniae*, *Chlamydia pneumoniae*

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 20-Jun-2000

C:Accession: EB1573

R:Red, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39.

A:Reference number: AB1500; MUID:20150255; PMID:10684935

A:Accession: EB1573

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-925 (REA>)

A:Cross-references: GB:AE002208; GB:AE002161; NID:G7189387; PIDN:AA38302.1; PID:G7189383

A:Experimental source: strain AR39, HL cells

C:Genetics:

A:Gene: CP0465

C:Superfamily: *Chlamydia trachomatis* hypothetical protein CT234

[illegible]

```

QY 234 SGVLVMSYWMINKV--LTDR--FYNPEMOKGKAKPRNMKDSFLYLDSPYL- 288
DB 192 FPIV-----FVYSKSLKSDHDFID-----THPPSLSAKLCEYDKTFLL 238
QY 289 ----LTLVLVAYGICINLIEVTKSQKLOYPNNMDS--EFMGNFSFMTGVSVLIML 342
DB 239 CFYFLMQLATA-----TEFNVIKIFEIOPASKEEFELVAHIKCSLMTSLGMCFAL 291
QY 343 FVGNVIRKFGMLTGAALTVPVWLLTGVFPALVIFRNQASGLVAMFCTTPLMAVVGA 402
DB 292 FAYSRIVRLG-----VNNIILFAPLCFLS-----LFLFWTFKTLTSLAVLA 333
QY 403 IONLSKSTKXALPDSTKEMAY--IPLDQEQVKKAKADIVVAARFGSGGALLQGGLLV 460
DB 334 M--VREBVTYALDNNIQLLYGP---NKIRNQ--IRIVSESFIPGLVMS---L 382
QY 461 ICGSIGAMTPYLAIVLFLIAMIWVSATKLNKLFPAQALK 501
DB 383 VC--FLSSQYVFCLIIISLIATILVC---LVRSYAKAKILK 418

```

RESULT 29

```

AF2383
hypothetical protein al14622 [imported] - Nostoc sp. (strain PCC 7120)
C/Species: Nostoc sp. PCC 7120
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C/Accession: AF2383
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasaemoto, S.; Matanabe, A.; Iriyuchi,
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A/Reference number: AB1807; MUID:21595285; PMID:11759840
A/Accession: AF2383
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1002 <KIR>
A/Cross-references: GB:BA000019; PIDN:BA076321.1; PID:G17133759; GSPDB:GN00179
A/Experimental source: strain PCC 7120
C/Genetics:
C/Gene: al14622
C/Superfamily: Chlamydia trachomatis hypothetical protein CT234

```

Query Match 5.7%; Score 151; DB 2; Length 1002;

Best Local Similarity 20.2%; Pred. No. 0.0013;

Matches 101; Conservative 88; Mismatches 196; Indels 114; Gaps 23;

```

QY 19 IHTHELKKVLPMPFMFCITFNVTYVLRDXTDLIVGAGSGAEALPFIKFWLVPCALIF 78
DB 24 LRPEBGEPTMMFAFTTVSVGLKRAEDSTVALFLDEYGAG--PLP---MMYTASAVMG 77
QY 79 M---LIYAKLSNIISKQALFYAVGTPFLIFPALPFTVIYPLRDVLHPTFADRLQAILPP 135
DB 78 MALVFVYSWMLQKIPFLR-----WVVAIAPCMIVPL-----ILLVLLRW 116
QY 136 GL-----LGLVALIRMTTPAIFYLALNGSVMLSLMFGPABNITIKHAKRFPALFG 189
DB 117 GIDVAFSVIVVFLRLWV--DSIYVNDLNTSIV-----ANQLEFIRIKITYPILS 167
QY 190 IGANISLASGRAIYMA---SKLRASVSEGVDPWGISLRLM---AMTYISGLVMSY 242
DB 168 SGLVADVDSGFSPLMLLEPAKN-----RVIMACGVMTISSALICLYST 213
QY 243 WINKNVLTDRFYNPEMOKGKAKPRNMKDSFLYLDSPYLTLTLVLVAYGICIN 302
DB 214 QYRTSFPFAPQRLIPOEQASRRRIQAP-----LKR--YTLQFAFVGLQIIGL 261
QY 303 LIETWKSQKLOYPNNMDS--YSEFMGNFSFMTGVSVLIMLFGVGNVIRKFGMLTGAALV 360
DB 262 LVDPQYLOELKI---NIGDRELAGFLGFGIIVGICELTTQMFVSSRLIERFGVFFPAL 318
QY 361 TPVWVLLTGIYFPAIVIFRNQASGLVAMFCTTPLMAVVGAL-IONLSKSTKXALPDST 419

```

```

DB 319 LPVAV---GFVYPMIVVLLRLGLOS-----LAFPMGLVLRKPFDELARTFVISS 367
QY 420 KEMAYIPLDQEQVKKAKADIVVAARFGSGGALLQGGLLVCSISGAMTPYLAIVLFT 479
DB 368 GPLLYQPI--PERISRTOV-----LSGGT---AEALATGTAG---IIIVITLFLV 409
QY 480 IAIWVSATKLNKLFPAQS 498
DB 410 CGLF-VEPATMOKWFISET 427

```

RESULT 30

```

B82219
transporter, BCC family VC1279 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C/Accession: B82219
R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
chardson, D.; Ermlaeva, M.D.; Vamathevan, J.; Basu, S.; Qin, H.; Dragoti, I.; Sellers, P
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A/Reference number: A82035; MUID:20406833; PMID:10952301
A/Accession: B82219
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-540 <HEI>
A/Cross-references: GB:AE004207; GB:AE003852; NID:99655761; PIDN:AAF94438.1; GSPDB:GN001;
A/Experimental source: serogroup O1; strain N16961, biotype El Tor
C/Genetics:
A/Gene: VC1279
A/Map position: 1
C/Superfamily: Escherichia coli probable carnitine transport protein

```

Query Match 5.5%; Score 145; DB 2; Length 540;

Best Local Similarity 20.5%; Pred. No. 0.0018;

Matches 106; Conservative 76; Mismatches 187; Indels 148; Gaps 24;

```

QY 69 WLVPFCALITMIL-----YAKL-----SNLSKQALFYAVGTRP--LIFFALF 109
DB 79 WLFVMSGNLFLPLFCVGLVISPFGKIRGQAKADHSFLSLMLLFFAAGMGIGLFWVSA 138
QY 110 PTYVY-----PLRDVLHPTFADRLQAILPGL--LGLVALIRMTTPAIFYLALNGS 161
DB 139 EPVAYITGMWETPL-----GVEANTPQAQLAMATVHFHGLHPW---AIYGV 183
QY 162 VMLSLMF-----WGFANETIKIHEAKRFPALFGIGANISL-- 196
DB 184 VALSLAFVCYNKGLPLSMRSVFYPLGDRAMGAVHVDLAV--LATLFGLATSLGLGA 241
QY 197 -LASGRAIYMAKLRASVSEGVDPWGISLRLMANTYISGLVMSYWMINKNVLTDR 254
DB 242 QQAASGIHWFGEFS-----GIGLOIAY-ICVVGTLTSLSVRGID----- 281
QY 255 FYNPEMOKGKAKPRNMKDSFLYLDSPYLTLTLVLVAYGICINLIEVTKSQKLO 314
DB 282 -----GGVKYISINIMIVAF-----LALVAIVIGSVITLSTPTLMAVLQN 324
QY 315 QYPMNDYSE-----FMGNFSF-----WTGVSVLIMLFGVGNVIRKFGMLTGAALT 362
DB 325 LIPLSNHRGDEDEWFGGVWFVWAMWISSPVGMFIASVSGRTIRF--IYAVLLIP 382
QY 363 VMLVLTGI-VFPALVIFR--NQASGLVAMFCTTPLMA-----VVGAIIONLSSTK 412
DB 383 TLVTLVMSVFGGLAVDQVINEI--GVLGQGLTLDVSLAMPQEDSLFVGVLSTIAVLV 441
QY 413 YALPDSKEMAYIPLDQ-----EOKYVSKAIDVVAARFGSGGALLQGGLLVI 461
DB 442 LVFFITSSDSGLVDSITSGKLDSPVLRVWMAVLEGAAVVALMTGGTEAVQALQ-- 499
QY 462 CGSIGAMTPYLAIVLFLIAMIWVSATKLNKLFPAQS 498

```

Db 500 AGVISTALPTVILLMLCVSLM--KGLKTEPEFMSS 534

RESULT 31

Probable transmembrane transport protein NMA1483 [Imported] - *Neisseria meningitidis* (serotype 14) [NCBI]
 C:Species: *Neisseria meningitidis*
 C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: E81839
 R:Perkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
 ; Holroyd, S.; Jorgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
 Nature 404, 502-506, 2000
 A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
 A:Reference number: AB1775; MUID:20222556; PMID:10761919
 A:Accession: E81839
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-660 <PAR>
 A:Cross-references: GB:AL162756; GB:AL157959; NID:G7380091; PIDN:CAE84716.1; PID:G7380131
 A:Experimental source: serogroup A, strain Z2491
 C:Genetics:
 A:Gene: NMA1483
 A:Superfamily: choline transport protein betT

| | | | | |
|-----------------------|--------------|-------------------|-----------------|-------------|
| Query Match | 5.5%; | Score 144; | DB 2; | Length 660; |
| Best Local Similarity | 20.7%; | Pred. No. 0.0027; | | |
| Matches 127; | Conservative | 99; | Mismatches 171; | Indels 216; |
| | | | | Gaps 34 |

QY 3Z IMFECITFENTYLRDPTKOTLIVGARG-----SGEBAIPFIKF-WLVYPCALIFM----- 79

Db 7 LEFVVCJVV-----VLVITVPQOVQUMIDRKEVIEFEFWMFVYLIFFSIFLGPFLI 56

QY 80 LIYAKUSNI-----LSKQALFYAVGTGF-LIFPALPPTYVPLADVIHPTER 125

Db 57 LVSISLIGNIRLGRDEDEVPBEFGLSMLAMFLAAGMGVGLMFGEVAPLMIHYSDDITAGTPE 116

QY 126 ADRLOAILPBGILGLVAILRNTFPAAYVYLAELMGSVMLSLMFGEPAHEITKIHAKR-- 183

Db 117 HROQOALL-----HTVFHWGVHMSV---YGTILALAYGFR---YKLPLARSC 161

QY 184 FYVL-----FGIGANI-SILLASGRAIV-----WASKIRASVSE--GVDPMGISRL 227

Db 162 FYPLKEKISGRBGADIDIMALLATPFGIITTLGFGASQLAGLOETGWIENSFSVOVL 221

QY 228 MAMTIVSGVLMAASYMINKVLTDPFRFYPEMOKKKGAK--PKAMMKDSFYL---YL 281

Db 222 IIAAWSLAVSA-----ISGVCKGVKVLSEINLGLAFILFFVL 261

QY 282 DNSPYILLTLVIAYGICI-----NLIEVTWKSQKIQYNNMDYSEBPGNCFSE----- 311

Db 262 AAGPTVYLLS---AFGDNIGNYLGNLVLSFEKT-----VAYEREHKPWESWSTVLYWAM 312

QY 332 ---WTGVVASVILMLFVGNGVIRKFGMLGALVTPPMVLTGIVFPALVIFRNOQSGELVAM 388

Db 313 WCSMAPFVGFLIRKISKGRITREF--VGVULIIPG--LFGVLMF-----TV 354

QY 389 FGTTPLML-AVYVGAIONILSKSTYALFDSITKENAYIPLDQOKV----- 433

Db 355 FGNTAIWLNDGVAGVLEKMTSSPELLF---KFPNVLPLDELNSIVSLVISLFFVTS 411

QY 434 ---KKA-----IDVVAARFGSGGALLIOQGLLVIGS1GAM 468

Db 412 DSGIYVNNITSBDKLSAPRMOAVMVGVLMSAVAVLLMRSG-----LGN1QM 461

QY 469 T-----PYLAVILFIITAIWL-VSA-----TKLN-----KLFLAQSALEQ 503

Db 462 TLIVSLPFLALMLIMFGFLMKGLSADKKCYFERIRVNPVSVPFTGKKMERLVQ1MSQ7QB 521

QY 504 EV---AOESAPA 513

Db 522 DILKPLKHTASPA 514

RESULT 32
E81101

transporter, BCGT family NMB1277 [imported] - *Neisseria meningitidis* (strain MCS8 serogroup C); Species: *Neisseria meningitidis*
C|Date: 31-Mar-2000 #sequence revision 31-Mar-2000 #text change 19-Jan-2001
A|Accession: B81101
R|Jettelein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.T.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Rhee, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzi, M. Science 287, 1809-1815, 2000
A|Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Vett
A|Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MCS8.
A|Reference number: A81000; MUID:2015755; PMID:10710307
A|Accession: B81101
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-675 <TEXT>
A|Cross-references: GB:AE002476; GB:AE002098; NID:g7226514; PIDN:AAFA1653.1; PID:g7226514
A|Experimental source: serogroup B, strain MCS8
C|Genetics:
A|Gene: NMB1277
C|Superfamily: choline transport protein belt

| | | | | |
|-----------------------|-------------------|-------------------|-------------|-------------|
| Query Match | 5.5%; | Score 144; | DB 2; | Length 675; |
| Best Local Similarity | 20.7%; | Pred. No. 0.0028; | | |
| Matches 127; | Conservative 100; | Mismatches 170; | Indels 216; | Gaps 344 |

```

QY 32 LMFEICIFENTVLRDXTDULIVGAPG-----SGAEIPIPKR-MUVPCALIFM----- 79
Db 22 LEFVCVUV-----VLVTVPDOVCKMIDRAKEVITFTEBSWFVYLTFSIFLGLLI 71
QY 80 LIVAKLSNI-----LSKQALFYAVGTPF-LIFPALFPYIVYLRDLVHPTFE 125
Db 72 LVSLSLGNIRLGRDEDEVPEFQPLSWLMLPFAAGGVGLMFFGVAEPLMHYFSDITAGTPE 131
QY 126 ADRLQAILPFGGLGLVALIRNWTTPAIFYVLAELMGSMVLSLMWGVPANETTKIHEAKR-- 183
Db 132 HROQOALL-----HTVFHMGVHAMSV--YGTIALALAFGRF--YKPLPLRLSC 176
QY 184 FYAL-----FGIGANI-SLASGRAIV-----WASKLRASVSE--GVDPMGISLRL 227
Db 177 FYPLIKKISGRQEDADIDIMALLATPFGIITTLGFRQASQLAGLOEMGMAENSFSVOVL 236
QY 228 MAMTIVSGVLVMSYWMINKVLTDRPFYNPEEMOKKQAK--PKAMKDSFL-----YL 281
Db 237 IIAVMSLAVSA-----ISGVKGVKVILSEINLGLALFLFPVL 276
QY 282 DRSYIILLFLVIAYGICI-----NUIEVTWKSQLKQYPMNNDYSEPMGNSF----- 331
Db 277 AAGSTVYLLS-----AFEDNIGNYLGMLVRLSFTK-----YAYERREKHPMESWTIVYMAW 337
QY 332 ---WTGVSVYLIMFVGANVIARKFGLMTGLALVPVNWLLTGIVFAFLVIFRNQASGLVAM 388
Db 328 WGSNAPFVGLEFIARISGRITREF--VFGVLLIPG--LFGVLMF-----TV 369
QY 389 FGTTPLML-AVVVGAIONILSKSTKYALFDSYTEMAYIPIDQOKY----- 433
Db 370 FGNTAIWLINDGVAGGMLEKMTSSPETLLF--KFENVYLPPELTSIVLSLIVISLFPVTS 426
QY 434 -----KQKAA-----IDVVAARFGKSGGALLQGGILVIGSGIGAM 468
Db 427 DSGIYVLIANTTSRDKGSLAPRQOAVMMGVLMASVAVALMRSGG-----LGNIQSM 476
QY 469 T-----PYLAVIILLFLIAIWL-VSA-----TKLN-----KLFLAQSAALKRO 503
Db 477 TLIVSLPFPALLIMCPSLWKGLSADKKYFETRVNPTSVFMTGQKWKERLVQIMSQTOBO 536
QY 504 EV---AQEDSAPA 513
Db 537 DILKFLKQIASPA 549

```

RESULT 33

B83719
multidrug resistance protein (efflux transporter) BHO54 [imported] - Bacillus halodurans
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 06-Jan-2003
C:Accession: B83719
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Maui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; M0ID:20512582; PMID:11056132
A:Accession: B83719
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-474 <STO>
A:Cross-references: GB:AF001508; GB:BA000004; NID:g10172890; PIDN:BA804273.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BHO54
C:Superfamily: multidrug-efflux transporter

Query Match 5.3%; Score 140; DB 2; Length 474;
Best Local Similarity 22.7%; Pred. No. 0.0039;
Matches 110; Conservative 74; Mismatches 212; Indels 88; Gaps 22;

QY 67 KFWLVPACILIMLYAKLSNLSKQAL-----FYAVGTPPII-PPALPPIVYPLR 117
DB 6 KKMVVV-CAVLFSGTMIINSMNPAIPQLMNVBADAVATGWTITTFMAMGMTPIRT 64
QY 118 DVLHPTPADRLQALIPGLGLVAII--RNMTFAAFVY--LAEIWSVM---LSIMF 168
DB 65 GVL-GDKIGKKAIVILGIFVLGSLGALSNLPSLIVFRLQIGIGGVMVPLSMITLIF 123
QY 169 -----WGFANEITK-----IHAKEFPYALFGIANISLASRAIVW 205
DB 124 DAFPRNERGLATGVGVASVMAPIPTLGGFIVTSNMKYLELVNIPFGLGIIAAVY 183
QY 206 ASKLRASVSEGVDPWGISL-----RLMAKTVISGLVIMASYMINKNKVLTDPFRNPE 259
DB 184 LPKIRASQIKIDRNGFLFVTAGVGSILAFGRMDLTHLT- WINGVLLILGLCLV 241
QY 260 EMOKGKGAAPKPMNKDSFLYDRSPYLLLTLLVIAVGI-CINLIEVTKSOLKQYRN 318
DB 242 FVQVEKRAEQPLD-----LSLFRIP--AVSLSIWVAGISSIGFAGIFLVPILLQ--Q 291
QY 319 MNDYSEPMGNBSFMTGVSVLIMLVGVGVIRKFGMLGALVTPPMV-----L 366
DB 232 VYDYPIMTGLVFLPSALFTGLTMSIGRMIDKRG-PSGIMTAGMMAVGTALGYLHL 350
QY 367 LTGIYF-FALVIFRQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDTKEMAYI 425
DB 351 ETGLMYITAFMAAIRGVGGLTTPATTGMAIIPGGLISRGSAAMNVLRQMSAAGIYFI 410
QY 426 PLDQKVKYKRAIDVVAARFGSGGALIQGLLVICSGIGAMTYLAIVILLFIATLV 485
DB 411 SVFFEVF-RQGLAL--VSTSFPEATLQAINEGFF-----VGFLLT-----ALSIPAYWLE 458
QY 486 SATK 489
DB 459 KKA 462

RESULT 34

T07946
cytochrome-c oxidase (EC 1.9.3.1) chain I - Cryptococcidium cohnii mitochondrion
C:Species: mitochondrion Cryptococcidium cohnii
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C:Accession: T07946
R:Norman, J.E.; Gray, M.W.
FEBS Lett. 413, 333-338, 1997
A:Title: The cytochrome oxidase subunit I gene (coxI) from the dinoflagellate, Cryptococ
A:Reference number: Z16237; M0ID:97424386; PMID:9280308
A:Accession: T07946
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-536 <NOR>
A:Cross-references: EMBL:AF012554; NID:g2358288; PIDN:AAB69658.1; PID:g2358289
A:Experimental source: strain WH-d
C:Genetics:
A:Gene: coxI
A:Genome: mitochondrion
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C:Keywords: cytochrome-c oxidase chain I homology <COI>
F:22-479/Domain: cytochrome-c oxidase chain I homology <COI>
F:75,395/Binding site: heme a iron (His) (axial ligands) #status predicted
F:255,304/Binding site: copper (His) #status predicted
F:255-259/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F:259/Binding site: oxygen (Tyr) #status predicted
F:385/Binding site: magnesium (His) (shared with chain II) #status predicted

Query Match 5.2%; Score 137; DB 2; Length 536;
Best Local Similarity 20.3%; Pred. No. 0.0075;
Matches 104; Conservative 73; Mismatches 174; Indels 162; Gaps 22;

QY 68 FWLVPACILIMLYAKLSNLSKQALFYAVGTPPII-PPALPPIVYPLRVLHPTPAD 127
DB 81 FYLVMP-----GLFGGNYFIVFQSPFVVP----- 109
QY 128 RLQ-AIIPGLGLVAII-----RNMTF-----AAFYVA-ELIWSVMSLMEWG 170
DB 110 RVNPFSLILSLTLLILSLISFEGGTGWTLYPLPSTSMNLSPPSIGIITGLLISG 169
QY 171 FANEITKHEKRPYALFGIANISLASRAIVWASKLRASVSEGVDPWGI---SLRL 227
DB 170 ISSVLTSLNFWITLISRSIGITLTKTS-----LFPWLLITSGMLL 211
QY 228 MAMTVISGLVIMASYMINKVLT--TDPFRNPEMOKGKGAAPKPMNKDSFLDPSY 286
DB 212 LTLPLTGLALMT-LSDINVTLPFDPF-----GGDPIYQHLFWFGHPEVY 259
QY 287 ILLP-----TLVIAVGI-----CINLI-EVTKSOLKQYRNNDYSEF 325
DB 260 ILIIPAGIISIIISGLIÖKIIPGNPSMIPMSISILGSVWGHMHTYIGLEDTDSYF 319
QY 326 MGNFSFWTGVSVLIMLVGVGVIRKFGMLTGAIVTPVMTL-----GIVFALVYFR 379
DB 320 SG-----VTIISLPTGFKI--FNWLSLYGNPSLLILTKNSSLFGLFL-LMFTI 367
QY 380 NQASGLVAMFGTTPMLAV-----VGAIONILSKSTKVALFDTKEMAYIPDQEKVK 434
DB 368 GGSYGII--IGNAAVDGLHDTYITIAHFHFLSLGAVIAIFSG-----ITFNEIKIG 419
QY 435 GKAIIDVVAARFGSGGALIQGLLVICG-----SIG 466
DB 420 SKNLPSCSSNNSRYNLVLTPIGILITRPGPHFGLFVMPRRRIDPFDSFISNPLSIG 479
QY 467 AMTPYLAIVILLFIATLVLSATKINKLFLAOSA 499
DB 480 SGITLSFGFLFKENCWITSIAQCSRLFLGHVA 512

RESULT 35

T14236
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Synechococcus sp. (strain PCC 7020
C:Species: Synechococcus sp.
A:Variety: strain PCC 7002
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 03-Jun-2002
C:Accession: T14236
R:Klughammer, B.; Sultemeyer, D.; Badger, M.R.; Price, G.D.
submitted to the EMBL Data Library, April 1997
A:Description: Involvement of ndh3, ndh3 and ORF427 genes in high affinity CO2 uptake
A:Reference number: Z17936
A:Accession: T14236
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-498 <KLU>
A:Cross-references: EMBL:U97516; NID:g2232044; PID:g2232047; PIDN:AAB62185.1
A:Experimental source: strain PCC 7002

C:Genetics:
A:Note: nhd3
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 4
C:Keywords: membrane-associated complex; NAD; oxidoreductase

Query Match 5.2%; Score 136.5; DB 2; Length 498;

Best Local Similarity 20.6%; Pred. No. 0.0076;

Matches 92; Conservative 60; Mismatches 139; Indels 155; Gaps 20;

```
QY 82 YAKLSNIIISKALFAVGT-----PFLIFPALFPTV-IYPLRDVILHPTFADRLQALIP 135
DB 55 YTEFHNNWISIIIGLNINLGVDSISPLIYINSILTLVAIYSIGESNHRKLYSLILINS 114
QY 136 GLIGVALIRMTFAFYVLAELMGSVWLSLMEFANETIKIHAKRFYALFGANIS 195
DB 115 GITG-ALIAN-----NLLFLFLFY---EL---ELIFYLLIAL----- 145
QY 196 ILASGRAIVMASKLRAVSEGVDPWGISRLIMAMTIVSGVLMAVW---WINKVVLTD 252
DB 146 -----W-----GGEKKGVASTKFLIYTAISGLVLAHFGIYWLQSSNPD 186
QY 253 PRFINPEEMQKGGKAKPKMMKDSFLYLDSPYLLITLIVAGICINLIEV-TWKSQ 311
DB 187 FENLTLEMLBENTK-----VILLITLIGGIRKIPVPLHTW--- 223
QY 312 LKLOYPMNDYSEFMGNFSFWTGVSVYLIMLFVGNVIRKFGWLTGALVTPVMTLIGIV 371
DB 224 LPDAIVEANP-----AVTVL-----GGVPAKLG-----TYGLV 252
QY 372 FFALVIFR---NOASGLVAMFGTTPMLAVVGAIONILSKTKYALPDSTKEMAYIPLD 428
DB 253 RFGIQLFPDWSVTSPALAVIGTYSVMYGLAALAQRLKMWAYS---SIGHMGIYIVS 309
QY 429 Q-----EOKYKKAIDVVAARFGKSGGALLIQ 456
DB 310 TAAGETELSLGVAQMISHSLIALLFHVLGIERKV-GRDLDVLNGLAMPVRLPLTS 368
QY 457 GLIVICSGIGAMTP---YLAVILLF 478
DB 369 SLTILGASAGITGVLGVFAEFLVF 394
```

RESULT 36

C83615 Probable MRS transporter PA0241 (imported) - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: C83615

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bt

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laidig, K.; Lim,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: C83615

A:Status: preliminary

A:Molecule type: DNA

A:Releases: 1-441 <STO>

A:Cross-references: GB:AE004462; GB:AE004091; NID:g9946077; PIDN:AG03630.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA0241

Query Match 5.1%; Score 135; DB 2; Length 441;

Best Local Similarity 21.1%; Pred. No. 0.0086;

Matches 110; Conservative 58; Mismatches 163; Indels 190; Gaps 26;

```
QY 32 IMFPCIFNTVAVLDYKOTLVG-----AGSGAEATPF--IKFWLVVPCALIFMLIYA 83
DB 39 LVFAATLNTYI-----DRAALGVMOPIIAEKMSWTAMDYANINFWQVGVYAIFFITLQGR 92
QY 84 KLSNLSKQALFYAVGTFPLIFPALFPTVIVPLRDVILHPTFADRLQALIPGGLGVAVI 143
```

DB 93 FIDKVGKRAFFLAV-----LLMSLATGAHGLATS 122

QY 144 LRMTFAAFYV-LAEI--WGSVMSLMPFGFANEITIHAKRYA--LFGICANISILA 198

DB 123 AAGFMVGRFIIIGLEAANYPACVKTITLMEPAGE-----RAVAAGIFAGTNGVAMV 174

QY 199 SGRAI-----VMASKLAASVSEGVDPWGISRLIMAMTIVSGVLMAVWINKVVLTD 253

DB 175 TPALLPILIGWG-----MOAFLCMSAL--GLWVLV-FWVRN----- 209

QY 254 RFYNPEEMQKGGK-----AKPKMMKDSFLYLDSPYLLITLIVIA----- 296

DB 210 -YNNPEEHPRVKQSELEFYIQQSEEPATRVFQSLIRRGWAPALAVSIAPVFWFYLY 268

QY 297 -----YGCINLIEVTKWSQLKQYPMNDYSEFMGNFSFWTGVSVYLIML----- 342

DB 269 WLPEFLNQYGLGISTVQM-----GIPLLIIMLTDADF 301

QY 343 FVGGNVIRKFGWLTGALVTPVMTLIGIVFPAL-----VIPRNASGL-VAMFGTTPMLA 397

DB 302 SVGGGILS--SWLIGRGPAPARLSMLLFACTIVGVFPANNSGLMIAV-----LAIA 354

QY 398 VVVGAIQ-----NLSKTKYALPDSTKEMAYIPLDOEKYKKAIDVVAARFG-----KS 449

DB 355 LAVGAHQAMTANIS-----LYMDYTPHGLMSTVFG-----FGMCAL 393

QY 450 GGALIQGLLVTCGSGIAMPYLAIVILLFTIATWVSATKL 490

DB 394 GGMFMTO---IVGGVLTATNNVAVLFTMIAMVFIATLWL 431

RESULT 37

E75212 Oligosaccharyl transferase PAB2202 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi

C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: E75212

R:anonymous, Genoscope

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

A:Reference number: A75001

A:Accession: E75212

A:Status: preliminary

A:Molecule type: DNA

A:Releases: 1-976 <KAW>

A:Cross-references: GB:AJ248283; GB:AL096836; NID:95457433; PIDN:CB49148.1; PID:9545765

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB2202

Query Match 5.1%; Score 134; DB 2; Length 976;

Best Local Similarity 20.9%; Pred. No. 0.025;

Matches 107; Conservative 79; Mismatches 157; Indels 168; Gaps 29;

```
QY 59 GAELIPRIRKMWVPCALIFMLIYA-----KLSNLSKQAL--F 95
DB 19 GKSILYPLK-RILPLAVIGFIVAYYLRHLTAGKFPDDPTFYHFIYLVLEKGLPKY 77
QY 96 Y-----AVGTPFLIFPALFPTVIVPLRDVILHPTFADRLQALIP--GLLGLIVA 142
DB 78 YPMAEAPFGSLIGEPGLY--ILPAITYKVSVPGNHF-----QAFLMWPPFVGLGVIA 131
QY 143 -----ILRNWTFAAFYVLAELMGSVWLSLMEFANETIKIHAKRYA--LFGIGA 192
DB 132 VYLLGRVYLNEM-----AGLMAVILSVSTANFRTSGNARGDGPMTLFLFSLVA 183
QY 193 NISLASSG--RAIVMS--KLRAVSEGV--DPMGISRLIMAMTIVSGVLMAVW 244
DB 184 MLVYLKENDIKKSLMWAVVLASISLGMWNGSPFLMW-----LGFASFQT 232
QY 245 INKNVLTDPFRFYNPEEMQKGGKAKPKMMKDSFLYLDSPYLLITLIVAGICINLI 304
DB 233 IALFI-----FGKIKELKRVKFRYP-----AYALAIAGGLTI--- 267
```

QY 305 EWTWKSQKLQYPMNNDYSEFMGNFSF--WTGVSVLIMLFVGNVIRKFGMLTALVTP 362
Db 268 -----PGIAKIGGFI-KPAFEVFLGLVLLVTIMLXGKFLVSDKKRFAVA 314
QY 363 VVWLLTGVFALVFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALPDSTREM 422
Db 315 VIVILG---FA-----GAYAVG--PKLFLMGGAAYO----STQ--VVOVQEL 352
QY 423 AVIPLDQOKVKGKAIIDVVAARFGSGS-----ALIQGLLVTC-----SIGMT 469
Db 353 A-----KTLISDIKLYGVGNGVLVFLFISIPFLIILGLYIMLMLKKSSESN 400
QY 470 PYLAVILLFIATIMVSAATKMLKFLAQSAL 500
Db 401 EYMLSVFYIMSLYLISLVA-VAFPLASTYAI 430
RESULT 38
D64137
best protein homolog - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 24-Sep-1999
C:Accession: D64137
R:Flischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M. Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: D64137
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-669 <TIG>
A:Cross-references: GB:U2843; GB:I42023; NID:91574554; PIDN:AAC3352.1; PID:91574560; T C:Superfamily: choline transport protein best
Query Match 5.1%; Score 133; DB 2; Length 669;
Best Local Similarity 19.8%; Pred. No. 0.019;
Matches 116; Conservative 87; Mismatches 167; Indels 216; Gaps 30;
QY 15 FLMPHTHEL-----KKVLPMPMPFCITFNTVTRDKTLIVAGPQGAARPIKFM 69
Db 27 FIAPBQTALNQASGLFANFSWTVLTSV-----FLGFL 63
QY 70 LVVPCAILFMIYAKLSNI-----LSKQALFYAVGTFP-LIPFALFTVTP 115
Db 64 LI-----LSVSLGNIKLGQDEEBEPFSLWMLFAAGMGVGLMFGVAEPLTHY 115
QY 116 LRDVHPEPEPADRLQAILPGLGLVALLRMNTFAAFVYLAELMGSVLSLMFGFANEI 175
Db 116 LSDITAGAEHKQOQALL-----HTLFHWGHAMAV---YGTIALALAYFGFRKL 163
QY 176 T-----KIHEAKRFYA-----LFGI-----GANISILASG-RATVWASKL 209
Db 164 PLALRSCTVPLKRIKINGIGDAIDVMMALLATLFGIITTLFGSQQLAGLEIQIMISQN 223
QY 210 RASVSEGVDPWGISRLIMAMTIVSGVLVMSVYMWINKVLTDPFRFYNPEEMQKKGAK 269
Db 224 SFALQVGI-----VVMCLAVFSAISGV-----GKGVK 252
QY 270 --PKKMMDSF-----LYIDRSYIILLTLVLAIVG-ICINLIEVWKSQKLQYPMNNDY 322
Db 253 ILSEINILAFCLLFLVLSGPTLYLAFSDNINYSNLYQLSFKT-----YAYEGEH 307
QY 323 SEFMGNFSF-----WTGVSVLIMLFVGNVIRKFGMLTALVTPVWVLLGIVFA 374
Db 308 TSMISGTVLYRWAMCSNAPFVGLFIARISKRTIRF--IFGVIVP--SLRGILWF- 361
QY 375 LVIFRNQASGLVAMFGTTPMLM--AVVGAIONILSKSTKVALPDSTREMAVTP----- 427
Db 362 -----TVGNTAVMLNDGIAAGGLGEFIS-SPEILLF-----KFLNVLPLPITTG 405

QY 428 -----DOE-----QKVKGAIDVVAARFGSGS-A 452
Db 406 FVSLIVLLFFITSADSGIYVANNIARSDKSLASPMQALIMGTLMSVVALVMOGGELA 465
QY 453 LIQOGLLVTCGSLGAMTPYLAIVLLFIATW-LVS-----ATKLN 491
Db 466 NLQMTLIVA-----LPFALLMLVMOFSLMKGLIADKKVFSTKVN 505
RESULT 39
GKECNK
nitrite extrusion protein - Escherichia coli (strain K-12)
N:Alternate names: nitrite facilitator
C:Species: Escherichia coli
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 01-Mar-2002
C:Accession: S05239; D64869
R:Naji, S.; Nohno, T.; Saito, T.; Taniguchi, S. FEBS Lett. 252, 139-143, 1989
A:Title: The narX gene product participates in nitrate transport induced in Escherichia A:Reference number: S05239; MUID:6938707; PMID:2668029
A:Accession: S05239
A:Molecule type: DNA
A:Residues: 1-463 <NCU>
A:Cross-references: EMBL:X15996; NID:942089; PIDN:CAA34126.1; PID:942091
A:Experimental source: strain K-12
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co. A.; Rose, D.D.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12. A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: D64869
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-463 <BLAT>
A:Cross-references: GB:AE00220; GB:U00096; NID:91787467; PIDN:AAC74307.1; PID:91787475; A:Experimental source: strain K-12, substrain MG1655
C:Genetics: nark
A:Gene position: 27 min
A:Map position: 27 min
C:Function:
A:Description: promotes a rapid rate of anaerobic nitrate reduction and the continuous e. A:Note: induction by nitrate
C:Superfamily: nitrate transport protein nark
C:Keywords: nitrate assimilation; nitrate transport; transmembrane protein
F:44-60/Domain: transmembrane #status predicted <TM1>
F:74-90/Domain: transmembrane #status predicted <TM2>
F:107-123/Domain: transmembrane #status predicted <TM3>
F:132-148/Domain: transmembrane #status predicted <TM4>
F:180-196/Domain: transmembrane #status predicted <TM5>
F:218-234/Domain: transmembrane #status predicted <TM6>
F:256-272/Domain: transmembrane #status predicted <TM7>
F:320-336/Domain: transmembrane #status predicted <TM8>
F:348-364/Domain: transmembrane #status predicted <TM9>
F:405-421/Domain: transmembrane #status predicted <TM10>
F:436-452/Domain: transmembrane #status predicted <TM11>
Query Match 5.0%; Score 132.5; DB 1; Length 463;
Best Local Similarity 22.3%; Pred. No. 0.014;
Matches 108; Conservative 69; Mismatches 174; Indels 133; Gaps 28;
QY 69 MLVPCAILFMIYAKLSNI-LSKQALFYAVGPFLLFPAIVIPLDVLPHTPEAD 127
Db 39 WISVPCILLAFPCWMLFSAVAVNDPKGFNTDQMLTALPSVSGALRV--PYSF-- 94
QY 128 RLQAILPPLGLVAIIL--BNWTFAPFYVLAELMGVSLMFWGFANEITKIHAKRFY 185
Db 95 -----WPIFGGRWT--AFST-----GLIILICWMLGFA-----VQDSTPY 130
QY 186 ALFEGIANISLLASGRAIVWASKLRASVS-----EGVDPMGISLRLIMAM 230
Db 131 SVFTI-----ISLLGFAAGANASSW-ANISFFPFKQKOGALGILGNGVSVMLQVAP 186

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:07:42 / Search time 53 Seconds
(without alignments)
1542.343 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 2630

Sequence: 1 MTKTEKPEFKLRSFLMPH.....AQSALEQEVAGQSDAPASS 515

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

A.GeneSeq.190un03:*

- 1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Query Length | ID | Description |
|------------|--------|-------|--------------|----|-------------|
| 1 | 2630 | 100.0 | 515 | 21 | AAV90265 |
| 2 | 2617 | 99.5 | 524 | 20 | AAV34951 |
| 3 | 2141 | 81.4 | 529 | 20 | AAV37153 |
| 4 | 1330.5 | 50.6 | 523 | 21 | AAAG13017 |
| 5 | 1330.5 | 50.6 | 618 | 21 | AAAG13016 |
| 6 | 1330.5 | 50.6 | 655 | 21 | AAAG13015 |
| 7 | 1220.5 | 46.4 | 587 | 21 | AAV50990 |
| 8 | 1156.5 | 44.0 | 569 | 21 | AAV50991 |
| 9 | 1054 | 40.1 | 498 | 21 | AAV50992 |

| | | | | | |
|----|--------|------|------|----|-----------|
| 10 | 1025.5 | 39.0 | 540 | 22 | AAV74452 |
| 11 | 1025.5 | 39.0 | 551 | 22 | AAV52325 |
| 12 | 939.5 | 35.7 | 535 | 20 | AAV37729 |
| 13 | 162 | 6.2 | 538 | 22 | AAAG89863 |
| 14 | 157.5 | 6.0 | 907 | 20 | AAV34888 |
| 15 | 152 | 5.8 | 420 | 20 | AAV36963 |
| 16 | 144 | 5.5 | 660 | 21 | AAV74601 |
| 17 | 144 | 5.5 | 660 | 21 | AAV74602 |
| 18 | 144 | 5.5 | 675 | 23 | AAU91074 |
| 19 | 134 | 5.1 | 660 | 21 | AAV74600 |
| 20 | 134 | 5.1 | 675 | 24 | ABP78636 |
| 21 | 132.5 | 5.0 | 463 | 22 | AAU34551 |
| 22 | 129 | 4.9 | 462 | 22 | AAU34568 |
| 23 | 125.5 | 4.8 | 713 | 21 | ABJ19074 |
| 24 | 122.5 | 4.7 | 713 | 21 | AAV58582 |
| 25 | 120.5 | 4.6 | 395 | 22 | AAU38299 |
| 26 | 120.5 | 4.6 | 425 | 22 | AAU36578 |
| 27 | 119.5 | 4.5 | 472 | 24 | ABJ19068 |
| 28 | 119.5 | 4.5 | 800 | 23 | ABP40810 |
| 29 | 119 | 4.5 | 517 | 20 | AAV49633 |
| 30 | 118 | 4.5 | 466 | 22 | AAU38238 |
| 31 | 118 | 4.5 | 694 | 23 | ABB93711 |
| 32 | 117.5 | 4.5 | 430 | 22 | AAU34182 |
| 33 | 117.5 | 4.5 | 744 | 23 | ABB48954 |
| 34 | 117 | 4.4 | 493 | 23 | ABB48123 |
| 35 | 116.5 | 4.4 | 437 | 24 | AAU38247 |
| 36 | 116 | 4.4 | 635 | 22 | AAE33467 |
| 37 | 115.5 | 4.4 | 410 | 22 | AAAG90107 |
| 38 | 115.5 | 4.4 | 410 | 22 | AAAB78947 |
| 39 | 115 | 4.4 | 392 | 22 | AAE05852 |
| 40 | 115 | 4.4 | 708 | 22 | AAAG91270 |
| 41 | 115 | 4.4 | 403 | 22 | ABB71346 |
| 42 | 114 | 4.3 | 478 | 22 | ABP39792 |
| 43 | 113 | 4.3 | 606 | 20 | AAV19800 |
| 44 | 113 | 4.3 | 633 | 20 | AAV19799 |
| 45 | 113 | 4.3 | 1095 | 22 | AAU00931 |
| 46 | 113 | 4.3 | 1095 | 22 | AAU69878 |
| 47 | 113 | 4.3 | 1095 | 22 | AAU01233 |
| 48 | 113 | 4.3 | 1095 | 23 | ABB95338 |
| 49 | 113 | 4.3 | 1095 | 23 | ABU71769 |
| 50 | 113 | 4.3 | 1104 | 21 | AAV95437 |
| 51 | 113 | 4.3 | 1104 | 21 | ABG61818 |
| 52 | 113 | 4.3 | 1681 | 23 | AAU78573 |
| 53 | 112.5 | 4.3 | 527 | 24 | ABJ37097 |
| 54 | 112.5 | 4.3 | 535 | 24 | ABJ37081 |
| 55 | 112 | 4.3 | 446 | 22 | ABB54869 |
| 56 | 112 | 4.3 | 680 | 22 | ABB65799 |
| 57 | 112 | 4.3 | 680 | 22 | ABB66097 |
| 58 | 111.5 | 4.2 | 388 | 15 | AAE57032 |
| 59 | 111.5 | 4.2 | 471 | 24 | ABJ19193 |
| 60 | 111.5 | 4.2 | 1083 | 22 | AAAG3160 |
| 61 | 111 | 4.2 | 360 | 18 | AAW44079 |
| 62 | 111 | 4.2 | 360 | 18 | AAW22178 |
| 63 | 111 | 4.2 | 741 | 22 | AAAG0701 |
| 64 | 111 | 4.2 | 1095 | 22 | AAU69879 |
| 65 | 111 | 4.2 | 1095 | 22 | AAU01234 |
| 66 | 111 | 4.2 | 1095 | 23 | ABB53339 |
| 67 | 111 | 4.2 | 1095 | 24 | ABU71770 |
| 68 | 110.5 | 4.2 | 425 | 22 | AAU34874 |
| 69 | 110.5 | 4.2 | 425 | 22 | AAAG8958 |
| 70 | 110.5 | 4.2 | 443 | 22 | AAAG6775 |
| 71 | 110.5 | 4.2 | 476 | 22 | AAAG91987 |
| 72 | 110.5 | 4.2 | 498 | 22 | ABP77193 |
| 73 | 110.5 | 4.2 | 691 | 23 | ABP69062 |
| 74 | 110.5 | 4.2 | 697 | 23 | AAAB78983 |
| 75 | 110.5 | 4.2 | 462 | 23 | AAAB55338 |
| 76 | 110 | 4.2 | 927 | 23 | ABP74086 |
| 77 | 109.5 | 4.2 | 665 | 23 | AAE21150 |
| 78 | 109.5 | 4.2 | 691 | 23 | AAE22915 |
| 79 | 109 | 4.1 | 538 | 22 | ABB61067 |
| 80 | 109 | 4.1 | 551 | 22 | ABB67783 |
| 81 | 109 | 4.1 | 683 | 22 | AAU44669 |
| 82 | 108.5 | 4.1 | 473 | 20 | AAV36998 |

Chlamydia pneumoniae
Chlamydia pneumoniae
Chlamydia trachoma
C. glutamicum prote
Amino acid sequenc
Amino acid sequenc
Neisseria meningit
Neisseria meningit
Neisseria gonorrhoe
N. gonorrhoeae ami
E. coli cellular p
E. coli cellular p
Pathogen specific
Sorangium cellulos
Salmonella typhi c
Staphylococcus aur
Staphylococcus spec
Staphylococcus epi
Wheat hexose cari
Salmonella typhi c
Herbicidally activ
Staphylococcus aur
Listeria monocytog
Listeria monocytog
Salmonella typhi c
Haemophilus influ
C. glutamicum prote
C. glutamicum SRT
Pseudomonas stutze
C. glutamicum prote
Drosophila melanog
Staphylococcus epi
B. burgdorferi ant
B. burgdorferi ant
Prostate-tumour de
Human prostate CDN
P788P amino acid s
Human P788P protei
Prostate cancer as
Human calcium chan
Prostate cancer-as
Mouse voltage gate
OspHt rice phospho
High-affinity phos
Lactococcus lactis
Drosophila melanog
Drosophila melanog
Serotonin receptor
Pathogen specific
C. glutamicum prote
S. thermophilus exo
S. thermophilus exo
C. glutamicum prote
Human prostate CDN
P788P amino acid p
Human P788P polyom
Prostate cancer as
E. coli cellular p
E. coli growth and
Corynebacterium gl
C. glutamicum prote
Streptococcus poly
Human polypeptide
C. glutamicum SRT
Lactococcus lactis
Candida albicans e
Human 52991 protei
Human transporter
Drosophila melanog
Drosophila melanog
Propionibacterium
Chlamydia trachoma

83 108.5 4.1 485 24 ABP76898 N. gonorrhoeae ami
84 108.5 4.1 587 22 AAU34198 Staphylococcus aur
85 108.5 4.1 596 22 AAU33748 Staphylococcus aur
86 108.5 4.1 603 22 AAU37074 Staphylococcus aur
87 108.5 4.1 604 22 AAU36932 Staphylococcus aur
88 108 4.1 422 23 ABP30076 Streptococcus poly
89 108 4.1 451 22 AAU33583 Streptococcus aeri
90 108 4.1 477 22 ABP47922 Pseudomonas faec
91 108 4.1 868 21 AAU41508 Arabidopsis thaili
92 108 4.1 879 21 AAU41507 Arabidopsis thaili
93 108 4.1 890 21 AAU41506 Arabidopsis thaili
94 108 4.1 933 22 AAU36224 Pseudomonas aeri
95 107.5 4.1 413 23 ABP49725 Listeria monocy
96 107.5 4.1 436 23 ABP29086 Streptococcus poly
97 107.5 4.1 507 23 ABP48420 Listeria monocy
98 107 4.1 527 19 AAU98268 H. pylori GHPD 335
99 107 4.1 841 21 AAU11039 S. xylosum mpf pr
100 107 4.1 1043 22 AAU92893 C glutamicum prote

ALIGNMENTS

RESULT 1

AAU90265 standard; Protein; 515 AA.

XX AAY90265;

XX 22-SEP-2000 (first entry)

DE C. pneumoniae ATP/ADP translocase protein sequence.

XX ATP/ADP translocase; Chlamydia infection; diagnosis; therapy.

XX Chlamydia pneumoniae.

XX OS WO200039157-A1.

XX PD 06-JUL-2000.

XX PF 22-DEC-1999; 99WO-CA01224.

XX PR 28-DEC-1998; 98US-0114060.

XX PR 12-MAR-1999; 98US-0123967.

XX PR 30-JUN-1999; 99US-0141271.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PI Murdin AD, Oomen RP, Wang J, Dunn P;

XX DR WPI; 2000-452368/39.

XX DR N-PSDB; AAA30922.

XX PT Novel Chlamydia polynucleotides and polypeptides, useful for diagnosis,

XX PS prevention and treatment of Chlamydia infection in mammals -

XX PS Claim 16; Fig 1; 81pp; English.

XX This sequence represents the Chlamydia pneumoniae ATP/ADP translocase
CC of the invention. The protein, DNA encoding it, or a vaccine containing
CC the DNA or protein, are useful for diagnosing, preventing or treating
CC Chlamydia infection. The sequences can also be used in a method for
CC the detection of Chlamydia infection. Primers or probes derived from the
CC DNA sequence are useful in diagnostic tests for detecting Chlamydia
CC infection.

XX SQ Sequence 515 AA;

Query Match 100.0%; Score 2630; DB 21; Length 515;
Best Local Similarity 100.0%; Pred. No. 2.2e-282;
Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKTEKFGKLRSLPFIHTHELKVLPMFLMFCITFNTYVLRDTKDTLIVGAPGSGA 60
Db 1 MTKTEKFGKLRSLPFIHTHELKVLPMFLMFCITFNTYVLRDTKDTLIVGAPGSGA 60
QY 61 EAIPPIKFWLVPCAIIFMLIVAKLSNLSKQALFYAVGTPLIFPALFPVIYPLRVL 120
Db 61 EAIPPIKFWLVPCAIIFMLIVAKLSNLSKQALFYAVGTPLIFPALFPVIYPLRVL 120
QY 121 HTEFADRLQALPPGLIGVAIILNMTFPAFYVLAELWGSVLSIMFMGFANETKTHE 180
Db 121 HTEFADRLQALPPGLIGVAIILNMTFPAFYVLAELWGSVLSIMFMGFANETKTHE 180
QY 181 AKRFALFGIGANISILASGRAIIVASVSEGVDPWGISLRLNMTIVSGVLMA 240
Db 181 AKRFALFGIGANISILASGRAIIVASVSEGVDPWGISLRLNMTIVSGVLMA 240
QY 241 SYWVINKKVLTDPRFYNEEMQKGGAKPPKNNKDSFLYIDRSPIYLLTLVIAIGIC 300
Db 241 SYWVINKKVLTDPRFYNEEMQKGGAKPPKNNKDSFLYIDRSPIYLLTLVIAIGIC 300
QY 301 INLEIVTKKSQKIQYPMNDYSEFMGNFSFWTGVSVLIMLVGQNVIRKFGMLTGALV 360
Db 301 INLEIVTKKSQKIQYPMNDYSEFMGNFSFWTGVSVLIMLVGQNVIRKFGMLTGALV 360
QY 361 TPVWVLTGIVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTK 420
Db 361 TPVWVLTGIVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTK 420
QY 421 EMAYIPLDQEQKVGKKAIDVVAARFGKSGALLIQGLLVICSGISGANTPYLAIVLFI 480
Db 421 EMAYIPLDQEQKVGKKAIDVVAARFGKSGALLIQGLLVICSGISGANTPYLAIVLFI 480
QY 481 AIWVLSATKLNKFLAQSALKEQEVADSDAPASS 515
Db 481 AIWVLSATKLNKFLAQSALKEQEVADSDAPASS 515

RESULT 2

AAU34951 standard; Protein; 524 AA.

XX AAY34951;

XX 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae transport polypeptide.

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

XX sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

XX KW vaccine; neutralising epitope.

XX OS Chlamydia pneumoniae.

XX OS WO9927105-A2.

XX PN 03-JUN-1999.

XX PF 20-NOV-1998; 98WO-IB01890.

XX PR 04-NOV-1998; 98US-0107078.

XX PR 21-NOV-1997; 97FR-0014673.

XX PA (BEST) GENSET.

XX PI Griffiths R;

XX DR WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae
PT Page 871-872; Disclosure; 1912pp; English.
PS AAY34584-Y35879 represent the proteins encoded by all the open reading
XX frames.

CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.

XX Sequence 524 AA;

Query Match 99.5%; Score 2617; DB 20; Length 524;

Best Local Similarity 99.6%; Pred. No. 6, 2e-281;
 Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTKTEKPPGKRSFLMPLHTEHLKVLPMFMFCITFNVTYLRDTKTLIVGAGSGA 60
 DB 10 MTKTEKPPGKRSFLMPLHTEHLKVLPMFMFCITFNVTYLRDTKTLIVGAGSGA 69
 QY 61 EAIPIFKFVLVPCAIIFMLIYAKLSNLSKQALFYAVGTPLIFPFIYPLRDVL 120
 DB 70 EAIPIFKFVLVPCAIIFMLIYAKLSNLSKQALFYAVGTPLIFPFIYPLRDVL 129
 QY 121 HPTFPADRLQALIPGLGLVALIRNWTFAFYVLAELMGVWLSLMFGFANEITKIH 180
 DB 130 HPTFPADRLQALIPGLGLVALIRNWTFAFYVLAELMGVWLSLMFGFANEITKIH 189
 QY 181 AKRFYALFGIGANISILASGRAIYVASKIRASVSEGVDPWGISLRLLMAMTIVSGLVMA 240
 DB 190 AKRFYALFGIGANISILASGRAIYVASKIRASVSEGVDPWGISLRLLMAMTIVSGLVMA 249
 QY 241 SYWMINKNVLTDPRFYNPEEMOKGKGAAPKKNMDSFLYLDRSYITLLTLVAYGIC 300
 DB 250 SYWMINKNVLTDPRFYNPEEMOKGKGAAPKKNMDSFLYLDRSYITLLTLVAYGIC 309
 QY 301 INLIETVWKSQKLQYPMNDYSEFMGNFSFWTGVVSLIMLVGNAVIRKFGMLTGALV 360
 DB 310 INLIETVWKSQKLQYPMNDYSEFMGNFSFWTGVVSLIMLVGNAVIRKFGMLTGALV 369
 QY 361 TPVWVLGTGIVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONLSKTYALFDSTK 420
 DB 370 TPVWVLGTGIVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONLSKTYALFDSTK 429
 QY 421 EMAYIPLDQEQYKGAIDVVAARFGKSGGALIOGGLVTCGSI GANTPYLAVALFTI 480
 DB 430 EMAYIPLDQEQYKGAIDVVAARFGKSGGALIOGGLVTCGSI GANTPYLAVALFTI 489
 QY 481 AIMLVSATKLNKFLAQSALKEQEVADSDAPASS 515
 DB 490 AIMLVSATKLNKFLAQSALKEQEVADSDAPASS 524

RESULT 3

AAY37153 standard; Protein; 529 AA.

XX AAY37153;

DT 07-OCT-1999 (first entry)

XX Chlamydia trachomatis transport protein.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 XX paratrachoma; inclusion conjunctivitis; genital disease; perinephalitis;
 XX nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 XX Bartholinitis; pneumopathy; venereal lymphogranulomatosis.

OS Chlamydia trachomatis.

XX MO9928475-A2.

XX

PD 10-UTN-1999.
 XX 27-NOV-1998; 98WO-1B01939.
 PF 04-NOV-1998; 98US-0107077.
 PR 28-NOV-1997; 97FR-0015044.
 PR 17-DEC-1997; 97FR-0016034.
 XX (GIST) GENSET.
 PA Griffiths R;
 PI WPI: 1999-371125/31.
 DR Genome sequence of Chlamydia trachomatis
 PT Disclousure; Page 935-936; 1755pp; English.
 XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 XX of Chlamydia trachomatis (see AAY201425). The polypeptides can be used as
 XX vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 XX can also be used to control growth of the microorganism. Chlamydia
 XX trachomatis is responsible for a large number of diseases, e.g. eye
 XX diseases such as conventional trachoma, nonendemic trachoma,
 XX paratrachoma, and inclusion conjunctivitis; genital diseases such as
 XX nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 XX Bartholinitis, pneumopathy; pneumonia in breast feeding infants;
 XX and venereal lymphogranulomatosis. The polypeptides of the invention
 XX may be of use in treating these diseases.

SQ Sequence 529 AA;

Query Match 81.4%; Score 2141; DB 20; Length 529;

Best Local Similarity 77.9%; Pred. No. 4e-228;
 Matches 409; Conservative 52; Mismatches 50; Indels 14; Gaps 3;

QY 1 MTKTEKPPGKRSFLMPLHTEHLKVLPMFMFCITFNVTYLRDTKTLIVGAGSGA 60
 DB 2 MTKTEKPPGKRSFLMPLHTEHLKVLPMFMFCITFNVTYLRDTKTLIVGAGSGA 61
 QY 61 EAIPIFKFVLVPCAIIFMLIYAKLSNLSKQALFYAVGTPLIFPFIYPLRDVL 120
 DB 62 EAIPIFKFVLVPCAIIFMLIYAKLSNLSKQALFYAVGTPLIFPFIYPLRDVL 121
 QY 121 HPTFPADRLQALIPGLGLVALIRNWTFAFYVLAELMGVWLSLMFGFANEITKIH 180
 DB 122 HPTFPADRLQALIPGLGLVALIRNWTFAFYVLAELMGVWLSLMFGFANEITKIH 181
 QY 181 AKRFYALFGIGANISILASGRAIYVASKIRASVSEGVDPWGISLRLLMAMTIVSGLVMA 240
 DB 182 AKRFYALFGIGANISILASGRAIYVASKIRASVSEGVDPWGISLRLLMAMTIVSGLVMA 241
 QY 241 SYWMINKNVLTDPRFYNPEEMOKGKGAAPKKNMDSFLYLDRSYITLLTLVAYGIC 300
 DB 242 SYWMINKNVLTDPRFYNPEEMOKGKGAAPKKNMDSFLYLDRSYITLLTLVAYGIC 299
 QY 301 INLIETVWKSQKLQYPMNDYSEFMGNFSFWTGVVSLIMLVGNAVIRKFGMLTGALV 360
 DB 300 INLIETVWKSQKLQYPMNDYSEFMGNFSFWTGVVSLIMLVGNAVIRKFGMLTGALV 359
 QY 361 TPVWVLGTGIVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONLSKTYALFDSTK 420
 DB 360 TPVWVLGTGIVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONLSKTYALFDSTK 419
 QY 421 EMAYIPLDQEQYKGAIDVVAARFGKSGGALIOGGLVTCGSI GANTPYLAVALFTI 480
 DB 420 EMAYIPLDQEQYKGAIDVVAARFGKSGGALIOGGLVTCGSI GANTPYLAVALFTI 479
 QY 481 AIMLVSATKLNKFLAQSALKEQEVADSDAPASS 513
 DB 480 AIMLVSATKLNKFLAQSALKEQEVADSDAPASS 524

RESULT 4
ID AAG13017 standard; Protein; 523 AA.
XX AAG13017;
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 12353.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128214.
PR 08-APR-1999; 99US-0128734.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131448.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134270.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0138763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 26-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.


```
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154033.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155658.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158023.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161355.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 50.6%; Score 1330.5; DB 21; Length 523;

Best Local Similarity 50.9%; Pred. No. 3.2e-138;

Matches 260; Conservative 92; Mismatches 136; Indels 23; Gaps 5;

```
QY 16 LMPHTHEKLVLPNFPCTTNYTVLRDTPKDTLIGAPSSGAEALPFIKFLVAVCA 75
DB 7 IFGVETTLTKKIVPGLMEFFCIIPYITIRDKDVLVYAKSSAEIIPFLKTVNLPMA 66
QY 76 IIFMFIYAKLSNLSKOLFVAVGTPEFCIFPALPFTVIVPLDVAHPTEFADRLQAIIP 135
DB 67 ICFMFLYTKLSNLSKALFVIVPFIYFPGAFGFWIPLSNLHPEALADKLATLGP 126
QY 136 GLGLVAILRNNTFAAFYVLAELWGSVMSLNFMGFANEITKIHEAKRPYALFGIGANIS 195
DB 127 REMGPLAIRIWSFCIFVYMAELMGSSVVSVLFMGFPANQITTVDEAKKFFYPLFGIGANVA 186
```

```
QY 196 LLASGRAIVWASKLRASVSEGVDPMGISLRLMAMTIVSGLVMASSWYINKVLTDPRF 255
DB 187 LIFSGRVYKIFSNMRKRLGVDGMASVLSKAMMSIVGKGLACFLYWWNR----- 238
QY 256 YNPEMOKGKKGAKPKKMMKDSFLYDRSPYILLTLVAYICINLIEVTKSQKIQ 315
DB 239 YVPLPTRSKKKKVPQGTWIESLKFVSSPYIDLATLVAYGISINLVEVTWMSKLKAQ 298
QY 316 YPMNNDYSEPMGNFSFTGTIVSVILMLFVGANTIRKRGWLTGALVPMVWLIGVFEAL 375
DB 299 FSPSENEYSAPMGDFSTGTGATFTMML-LSQYVFKKYGWVAAKITPTVILLTGVAFFSL 357
QY 376 VIFPNQASGVAMFGTTPMLAVVGAIONILSKSTYKALFDS TKEMAYIPLDQEQRYKG 435
DB 358 IIFGPPAPVAKLGMTPPLAAVYVGAQNIFSKSAKSYLFDCKEMAYIPLDQDTYKG 417
QY 436 KAAIDVVAARFGKSGGALLIOGGLVIGSIGAMTPYLAIVLLPFIATWVSATKL---N 491
DB 418 KAAIDVVCNPLGKSGGALLIOGFMLIFGSLANSTPYGIVLLIGVTAMLAAKSLGQFN 477
QY 492 KLFLAQGALKQEQ-----VAQEDSAPA 513
DB 478 TLMSEELEREMERASSVKIPVVSQED-APS 507

RESULT 5
AAG13016
ID AAG13016 standard; Protein; 618 AA.
XX
XX AAG13016;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 12352.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
```

PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0144941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143524.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149910.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151393.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154019.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158359.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.


```
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 50.6%; Score 1330.5; DB 21; Length 655;
Best Local Similarity 50.9%; Pred. No. 4,4e-138;
Matches 260; Conservative 92; Mismatches 136; Indels 23; Gaps 5;

QY 16 LMPHTHELRKVLPMPFMFCITRNNTYVLRPTKOTLLVGBGSAEAIPIFKFVLVPCA 75
DB 139 IFGEVTVTKKIVPLGFMFCILFNNTYLRTDKVLVVTAKGSAEIIPIKTVNLPM 198
QY 76 IIFMLIYAKLSNISKOALFYAVGTPFLIPALPPTIYIPRDVHPTEFADRIQALIPP 135
DB 199 IGFMLIYTKLSNVSKALPFTTVLPPTIVYGAGFPWTPISNIIHEBALDKLATLGP 258
QY 136 GLLGIVAILRNWTPAAYVLAELMGVYMLSMFMFANEITKJHEAKRFYALFGIGANIS 195
DB 259 RFMGPLAIMRIMSPCLFYVAHELWGSVSVLFWGFANQITTVDEAKKFPPLFOLGANVA 318
QY 196 LLAGGRAIWAASKIRASVSEGVDPWGISLKLMAMTIVSGIYLMASITWINKVLTDPRF 255
DB 319 LIFSGRTVYKFFSNRNKVLGPGVDGMAVSLKAMMSIVGMGLAICFLVMWNR----- 370
QY 256 YNPEMOKKKGAKPKRNKMKDSFLYLDSPYILLTLTVIAYGICILIEYTWKSOXLQ 315
DB 371 VYLPPTSKKKKKVYPQGMTHESLKFVSSPYIRDLATLVVAIGISINLVEYTWKSKLKAQ 430
QY 316 YPNNDYSEFMGNESFMTGVSVLIMLFVGANVIRKEGMLTGAIVTEPMVLTGIVEPAL 375
DB 431 FPSPNBYSAFMGDRSTGTGATFTMML-LSQYVFKKYGWGAAKITPTVLLTGVAFESL 489
QY 376 VIFPNQASGLVAFGTTPPLMLAVVNGAIONILSKSTYALPDSITKEAAYIPLDQEQVKG 435
DB 490 ILFGPPAPLVAKLGMPPLAAVYVGAIONIFKSATYSLPDCKEKAAYIPLEDRTVKKG 549
QY 436 KKAIDVVAARFGSGGALIOOGLLVICSGANTPYLAVLTLFIATMLVSATK-----N 491
DB 550 KKAIDVVCNPLGSGGALLIOQFMILITFGSLANSTPIYGVILLGIVTAMLAASLBQFN 609
QY 492 KLFLAQSALKQOE-----VAQEDSAPA 513
DB 610 TLMSEELEREMERASVKKIPVVSQED-APS 639
```

| ID | AAVS0990 | standard; Protein; 587 AA. |
|----|---|---|
| AC | AAVS0990; | |
| DT | 10-MAR-2000 | (first entry) |
| DE | A. thaliana | AATP1 protein. |
| XX | ADP/ATP translocator; | transgenic plant; plastid; oil; starch; |
| XX | yield increase; AATP1. | |
| OS | Arabidopsis thaliana. | |
| PN | WO9958654-A2. | |
| PD | 18-NOV-1999. | |
| PF | 12-MAY-1999; | 99WO-EP03292. |
| XX | 13-MAY-1998; | 98DE-1021442. |
| XX | (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH. | |
| PA | (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN. | |
| XX | Neubaus E, Wohlmann T, Graeve-Kampfenkel K, Tjaden J, Schnell J; | |
| PI | Martini N; | |
| DR | WPI; 2000-062295/05. | |
| XX | Transgenic plants with modified plastid ADP-ATP translocator activity | |
| PT | - | |
| PS | Example 2; Fig 3; 60pp; German. | |
| XX | This invention describes a novel genetically modified transgenic plant | |
| CC | cell, where the genetic modification is the introduction of a foreign | |
| CC | nucleic acid molecule, which exists, or its expression leads, to increase | |
| CC | of plastid ADP/ATP translocation activity in particular compared to wild | |
| CC | type plant cells. Nucleic acid molecules encoding a plastid ADP/ATP | |
| CC | translocator can be used to create transgenic plants and plant | |
| CC | cells with increased yields, especially an increased oil and/or starch | |
| CC | content. The plants are able to synthesize a modified starch. Plants, | |
| CC | such as maize, with an increased oil and/or starch content are of | |
| CC | commercial interest especially for the wet milling industry. Increased | |
| CC | starch content is important in food crops and in the paper and textile | |
| CC | industries. This sequence represents the Arabidopsis thaliana AATP1 | |
| CC | protein which is used in the method of the invention. | |
| XX | Sequence 587 AA; | |
| XX | Query Match | 46.4%; Score 1220.5; DB 21; Length 587; |
| DB | Best Local Similarity | 51.7%; Pred. No. 6e-126; |
| DB | Matches 252; Conservative 76; Mismatches 144; Indels 15; Gaps 4 | |
| QY | 16 IWPITHHEIKVLPMEFMFCITFNTYVLRDXTLLIVGAPSGABAIPIKFWLVPCA | 75 |
| DB | 104 IFGVAVATLKITPIGLMFCILFNTYLRDXTDVLVVAKGSSABIIPIFLKTWVNLPM | 163 |
| QY | 76 IIFPMILYALSNILSKQALFYAVGPELIFPALFPYIVIPLDVDVLPTEPADRLQAILRP | 135 |
| DB | 164 IGFMLLYTLKSNVLSKKALFYIVIVPFIYFGSGFVMPYLSNYIPEALADKLTLTLP | 223 |
| QY | 136 GLUGVALIDRNNTFAFYVLAELMGVMSILMFPGFANETIKIHEAKRFYALFGIGANTIS | 195 |
| DB | 224 RMGSPALAIRINSFCLFYVMAELMGVVSVLPFGWGANQITTYDEAKKFTYPLEGIGANVA | 283 |
| QY | 196 LLASGRALYVASKLRASVEGVDPEWGISLRILMAMTIVSGVLVMASTYVWINKVLTDP | 255 |
| DB | 284 LIFSGLTYVYFENLRKNIAPGVD--GSFVESHDEHGGNGTRICLSIGGSNRVPLPTPS | 341 |

```

QY      YNPSEMQGKKGAKRKMMKXDSPLYLDRSPYTLTLTLVIAVGICINTLEVTWKSQLKQ 315
Db      342 KKXKE-----KPKMGTESLKFVSSPYIRDLATLVAVGISINLVEVTWKSRLKQ 393
QY      316 YPNNDVSEFPMGNSFMTGVVSLIMLFVGVGNVRKFGMLTGALTVPWVLLTGIVFPAL 375
Db      394 FPPSPREYSAFMGAFSTCTGVATFTMTL-LSQVFNKYGMGVAAKITPFTVLLLTGVAFFSL 452
QY      376 VIFNRQASGLVAMFETFTLMLAVVGAIONLSKSTKXALPSTEMAYIPLDQEKVKG 435
Db      453 ILFGGFPAPLVAUKGMTPLLAVVGAIONIFSKAQXSLFDPCKEMAYIPLDDETRVKG 512
QY      436 KAAIDVVARCKSGGALIOOGLVTCISGAMTPYLAVTLIFIAIWLVSATKL-----N 491
Db      513 KAAIDVVCNPLCKSGGALIQPFMLISFQSLANSTYLGMLLVITYAMLAALKSLBGQIN 572
QY      492 KLFLAQ 498
Db      573 SLRLKKS 579

RESULT 8
ID      AAY50991
XX      AAY50991 standard, Protein, 569 AA.
XX      AAY50991;
AC      AAY50991;
DT      10-MAR-2000 (first entry)
XX      A. thaliana AATP2 protein.
DE      A. thaliana AATP2 protein.
XX      ADP/ATP translocator; transgenic plant; plastid; oil; starch;
XX      yield increase; AATP2.
KM      Arabidopsis thaliana.
OS      Arabidopsis thaliana.
XX      Arabidopsis thaliana.
PN      WO958654-A2.
XX      18-NOV-1999.
PD      18-NOV-1999.
XX      12-MAY-1999; 99WO-EP03292.
PF      12-MAY-1999; 99WO-EP03292.
XX      13-MAY-1998; 98DE-1021442.
PR      13-MAY-1998; 98DE-1021442.
XX      (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
PA      (PLAN-) MAX PLANCK GBS FORDERUNG WISSENSCHAFTEN.
XX      Neuhaus E, Moehlmann T, Graeve-Kampfenkel K, Tjaden J, Schell J;
PI      Martin N;
XX      WPI; 2000-062295/05.
DR      WPI; 2000-062295/05.
XX      Transgenic plants with modified plastid ADP-ATP translocator activity
PT      -
XX      Example 2; Fig 3; 60pp; German.
PS      Example 2; Fig 3; 60pp; German.
XX      This invention describes a novel genetically modified transgenic plant
XX      cell, where the genetic modification is the introduction of a foreign
XX      nucleic acid molecule, which exists, or its expression leads, to increase
XX      of plastid ADP/ATP translocation activity in particular compared to wild
XX      type plant cells. Nucleic acid molecules encoding a plastid ADP/ATP
XX      translocator can be used to create transgenic plants and plant
XX      cells with increased yields, especially an increased oil and/or starch
XX      content. The plants are able to synthesize a modified starch. Plants,
XX      such as maize, with an increased oil and/or starch content are of
XX      commercial interest especially for the wet milling industry. Increased
XX      starch content is important in food crops and in the paper and textile
XX      industries. This sequence represents the Arabidopsis thaliana AATP2
XX      protein which is used in the method of the invention.
XX      Sequence 569 AA;

```


KM bronchitis; sinusitis; atherosclerosis; asthma; vaccine; antimicrobial;
 KM antibacterial; diagnosis.
 OS Chlamydia pneumoniae.
 XX WO200121803-A1.
 XX PN
 XX 29-MAR-2001.
 PD
 XX 15-SEP-2000; 2000WO-CA01087.
 XX PF
 XX 17-SEP-1999; 99US-0154326.
 XX PR
 XX (AVENTIS PASTEUR LTD.
 PA
 XX Murdin AD, Oomen RP, Wang J, Dunn P;
 PI
 XX WPI; 2001-316102/33.
 DR N-PSDB; AAF84486.
 XX
 XX
 PT New Npt2cp (ADP/ATP translocase) polypeptides and nucleic acids
 PT encoding the polypeptides useful for treating, preventing or diagnosing
 PT Chlamydia infections, particularly infections caused by Chlamydia
 PT pneumoniae
 PS Claim 16; Fig 1; 79pp; English.
 XX
 XX This sequence represents the Chlamydia pneumoniae Npt2cp (ADP/ATP
 CC translocase) protein. Chlamydia pneumoniae Npt2cp nucleotides and
 CC proteins may be used for preventing, treating or diagnosing
 CC Chlamydia infections, particularly those caused by Chlamydia
 CC pneumoniae. Chlamydia pneumoniae can be responsible for both upper and
 CC lower respiratory tract infections. It is the third most common cause of
 CC community acquired pneumonia, and can also cause bronchitis and
 CC sinusitis. It is also linked to diseases other than respiratory tract
 CC infections, being associated with atherosclerosis and asthma.
 CC Npt2cp nucleotides may be used for the recombinant production of the
 CC protein, as genetic vaccines, in the construction of vaccine vectors
 CC such as poxviruses, and in the construction of attenuated Chlamydia
 CC strains which can over-express an Npt2cp nucleotide, or which can
 CC express it in a non-toxic, mutated form. Npt2cp polypeptides can be used
 CC in vaccine compositions, and may also be used as diagnostic reagents for
 CC detecting the presence of anti-Chlamydia antibodies in a sample.
 XX
 SQ Sequence 540 AA;
 Query Match 39.0%; Score 1025.5; DB 22; Length 540;
 Best Local Similarity 39.4%; Pred. No. 2.3e-104;
 Matches 214; Conservative 107; Mismatches 185; Indels 37; Gaps 9;
 QY 1 MTKTEKPEFGKLRSLFMPHITHELKKVLPMLMFCITFNVTYVLRDTKDTLIVGAPGSGA 60
 DB 1 MOSSEVKPFSRLRAYICPIYKSBFSKFVPLFLAFVFGNYCLKMKMDTLIVGSDGA 60
 QY 61 EAIPIFKFPLVVPACATIFMLIYAKLSNLSKQALFAVAVGTPELIPALFPPTYIPPLDVL 120
 DB 61 EYIPIFKFVGVGAVITVWVWGLGSRPRDTVFCFMAALGFFELFAVITVFGDSL 120
 QY 121 HPTFPAQLQALIPGLGLVAILRNWTPAFAVLAELWGSYMLSMFWGANETIKTHE 180
 DB 121 HUNSLADKLQELPQGLRFGFIWVRYSYIYVMSLSVVLGMLFGLANQTTTITE 180
 QY 181 AGRFVALFSGIGANISILASGRALWASKR-ASVSGVDPW-GISRLMLAMTIYGVYL 238
 DB 181 AGRFVALINTGNLSSICAGEISYMMGKOTFVAISPACDSMWSVMLNLTMLTT-CSGLIM 239
 QY 239 MASVYMWINKV--LTPRFYNEEMOKGKGA-----KPKAMKDSFYLRSPYI 287
 DB 240 I-----WLYRRIHHTLTDTSIPSRRLAEGAAATANLKKKKPKAKARLFLHLQSKRL 295
 QY 288 LLLTLVLVAYGICINILEVTKSOLKQYNNNDVSEFNGNFSFWTGVSVLIMLFGVGN 347
 DB 296 LGIAITVLSYNVILHFEVVMKQVQVQIYSSHVEFNGWSRITTLIGVSVLAVALTNGQ 355

QY 348 VIRKGMLTGALVTPVWVLTNGIVFPALVI-FRNOASGLVMEFGTTPMLAVVGAIONI 406
 DB 356 CIRKGMVTVGALVPLVWLVSGLLFPGTIPAKKADISIFGVLGMPALAAATMGAMNV 415
 QY 407 LSKSTKVALPDSTEMAYIPLDOEKVKGAIDVVARFPGSGALIQGLLVICSGIG 466
 DB 416 LSRGKTFPFQGTQEMAFIPLSPEDKHKGAIDGVSRIGKSGSLIYQGLLVIFSSVA 475
 QY 467 AMTPYLAVILFLTAIMLVSAFKLKLFLAQA-----LK-----RQVAAQE 508
 DB 476 ASLVNVALVLLIIMVWVAVVAYIGKEYSRADAVALTKPKSPSSSVREAOESVQE 535
 QY 509 DSA 511
 DB 536 EWA 538
 RESULT 11
 ID AAY35235
 AAY35235 standard; Protein; 551 AA.
 AC AAY35235;
 XX
 DT 13-SEP-1999 (first entry)
 DE Chlamydia pneumoniae transport polypeptide.
 XX
 XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KM vaccine; neutralising epitope.
 OS Chlamydia pneumoniae.
 XX
 XX WO9927105-A2.
 EN
 XX 03-JUN-1999.
 PD
 XX 20-NOV-1998; 98WO-1B01890.
 PF
 XX 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GBST) GENSET.
 XX
 PI Griffiths R;
 DR WPI; 1999-357842/30.
 XX
 XX Genome sequence of Chlamydia pneumoniae
 PT
 XX Page 1080-1081; Disclosure; 1912pp; English.
 PS
 XX AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAY34584) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.
 CC
 SQ Sequence 551 AA;
 Query Match 39.0%; Score 1025.5; DB 20; Length 551;
 Best Local Similarity 39.4%; Pred. No. 2.3e-104;
 Matches 214; Conservative 107; Mismatches 185; Indels 37; Gaps 9;
 QY 1 MTKTEKPEFGKLRSLFMPHITHELKKVLPMLMFCITFNVTYVLRDTKDTLIVGAPGSGA 60
 DB 1 MOSSEVKPFSRLRAYICPIYKSBFSKFVPLFLAFVFGNYCLKMKMDTLIVGSDGA 60

```

Db      12 MOSSEVPSRGLRAVLCPIYKSEPSKVPLEFLAFVGVNCLLKNMKDTLVI VGS DGA 71
QY      61 EALPFIKFMVLCALIFMLIYAKLSNLSKQALFVAVGPFLIFPALFPTVYIPRLDV 120
Db      72 EVLPFLKVMIVGAVLVTVVWGLSGRPDRDVFYCFMAAFGFFFLPAVVIITPYGDSI 131
QY      121 HPTFADRLQALIPGGLGLVALIRMTFAFVYLAELMGSVMSLMFQFANEITKHE 180
Db      132 HMLSLDKQDELPOGRGFIVWVRVWSYIYVMSLSMSVLSLMLFQNLQITITE 191
QY      181 AKRFYALFGIGANISLIASGRAIWAASKLR-ASVSEGVDPW-GISRLIMMTVSGVLV 238
Db      192 AGRFYALINTGLMSSICAGEISYMMGKQTFVAVSPACDSMHSVMNLITLIT-CSGLIM 250
QY      239 MASVMMINKV--LTDRFVNPBEOMKGGKA-----KPKMMKOSFLYLDSPYI 287
Db      251 I---WLYRRIHHLITDTSIPPSRRVLAEGAAFTANLKEKKRKARAKMLFHLIQSRYL 306
QY      288 LLLTLVIAVAGICINILEVTWKSQALKQYNNMNDYSEFMGNFSPMTGVSVLIMLFGVN 347
Db      307 LGLAIVLSNLYIHLEFVVMKQVQSIYSSHYEPNGYSRITTLIGVSVLAVALITQ 366
QY      348 VIRKFMVLCALTPVAVVLTGTIVFPALVI-FNQASGLVAMEGTTPLMLAVVGAIQNI 406
Db      367 CIRKMGMTVGAIVTPLVMLVSGLLFGITIFAARDISIFGVLMPLALAAVTGMQVY 426
QY      407 LSKSTKVALPDSRKEMAYIPLDOEKVKGAIDVVAARFGKSGGLIOGGLVITGSGIG 466
Db      427 LSKGIVTFEFDQKEMAFIPLSPEDKNGKAIDGVVSRIGSGSLIYQGLVIFSSVA 486
QY      467 AMTPYLAIVLLFIATIMVSATKLNKFLAOSA-----LK-----EOEVAQ 508
Db      487 ASLNVALVLLIIMVMAVVAIIGKEYYSRAADAVALIKQKPESSIVREAGSVBE 546
QY      509 DSA 511
Db      547 EMA 549

RESULT 12
AAV37729
ID      AAV37729 standard; Protein; 535 AA.
AC      AAV37729;
XX
XX      07-OCT-1999 (first entry)
XX
XX      Chlamydia trachomatis transport protein.
XX
XX      Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
XX      paratrachoma; inclusion conjunctivitis; genital disease; perhepatitis;
XX      nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
XX      bartolinitis; pneumopathy; venereal lymphogranulomatosis.
XX
XX      Chlamydia trachomatis.
XX
XX      WO928475-A2.
XX
XX      10-JUN-1999.
XX
XX      27-NOV-1999; 98WO-1B01939.
XX
XX      04-NOV-1998; 98US-0107077.
XX      28-NOV-1997; 97FR-0015041.
XX      17-DEC-1997; 97FR-0016034.
XX
XX      (GSEST ) GENSEST.
XX
XX      Grifffals R;
XX
XX      MPI; 1999-371125/31.
XX
XX      Genome sequence of Chlamydia trachomatis
XX

```

```

XX      XX
PS      Disclosure; Page 1329-1330; 1755bp; English.
CC      CC
CC      AAY6754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC      of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
CC      vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC      can also be used to control growth of the microorganism. Chlamydia
CC      trachomatis is responsible for a large number of diseases, e.g. eye
CC      diseases such as conventional trachoma, nonendemic trachoma,
CC      paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC      nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
CC      perhepatitis, bartolinitis; pneumopathy in breast feeding infants;
CC      and venereal lymphogranulomatosis. The polypeptides of the invention
CC      may be of use in treating these diseases.
SQ      Sequence 535 AA;
QY      Query Match 35.7%; Score 939.5; DB 20; Length 535;
QY      Best Local Similarity 37.2%; Pred. No. 7.8e-95;
QY      Matches 195; Conservative 113; Mismatches 195; Indels 21; Gaps 6;

8 PPKIRSFMLPIHTELKVLPMFLMFCITFNTVTLRDYKOTLIVGAPSGAALPFIK 67
2 PFRSGDXFPFIYSEFSKFIPLFLAFVGVNVALKTQDSIVLVGSRAGAIVIFLK 61
68 FVLVPCALIFMLIYAKLSNLSKQALFYANGTPELIFPALFPTVYIPRLDVLPTEPAD 127
62 VMGIVPGAVITWYIYGMWSRRYSRGVTFISLVGFLFGLFATVITPIGALHLNKLA 121
128 RLQALIPGGLGLVALIRMTFAFVYLAELMGSVMSLMFQFANEITKHEAKRFYAL 187
122 KLQSLIPPGRGFVVMQVMSYSLYYMSLSMSVLSLTFWGVANHITSVREAGRYAL 181
188 FGIGANISLIASGRAIWAASKLR-ASVSEGVDPGIGISRLIMMTVSGVLVMASTYWIN 246
182 INTGLMSSVPAQGEVSLMGRSPVIAFPMAVDPMHEMLNITLIVLAGVITLYLQKD 241
247 KNVLTPRFVNP-----EMQKGGKAKPKPMNKDSFLYDRSPYITLLTLVIAV 297
242 R--LMDETSMLEELAAEMSYAQLKKEKSKAK--AKSLFALLRSRYTLGLIADVLSY 297
298 GICINLIEVTWKSQALKQYNNMNDYSEFMGNFSPMTGVSVLIMLFGVGNVIRKFWL 357
298 NLVILHLEFVVMKQVQSIYSSHYEPNGYSRITTLIGVSAAGIFAAQGTIRMGWTVG 357
358 ALVTPVAVVLTGTIVFP-ALVIFRNQASGLVAMEGTTPLMLAVVGAIQNLSTKVALP 416
358 ALVTPVAVVLTGTIVFP-ALVIFRNQASGLVAMEGTTPLMLAVVGAIQNLSTKVALP 417
417 DSTKEMAYIPLDOEKVKGAIDVVAARFGKSGGLIOGGLVITGSGIGAMPYLAIV 476
418 DQTKEMAFIPLSPEDKNGKAIDGVVSRIGSGSLIYQGLVIFSSVAASLNATITVL 477
477 LFIATIMVSATKLNKFLAOSA-----LKEQVVAQEDSAPAS 514
478 LLAGSWIFVIAMVLAEGAAFTANLKEKKRKARAKMLFHLIQSRYL 521

RESULT 13
AAG89863
ID      AAG89863 standard; Protein; 538 AA.
AC      AAG89863;
XX
XX      26-SEP-2001 (first entry)
XX
XX      C glutamicum protein fragment SEQ ID NO: 3617.
XX
XX      Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX      organic acid synthesis.
XX
XX      Corynebacterium glutamicum.
XX

```


QY 115 PLRDLVHPTERADRLQALIPGGLGLVALIRNWTFAFYVLAELMGVMS-LMPWGFAN 173
 Db 110 -----AIDPP-----RSPLF--FYRIV-IMSLTILSTSPWGFVD 141
 QY 174 EITKHEAKRFYALPGIGANISLASGRAIVASGLRVSQVDPFGISRLMAMTIV 233
 Db 142 QFNLQDGRKHFCTN-----AIFLGDAI--GSGIISLVHTTIGIQLILFTALVLT 194
 QY 234 SGLVIMASYWVWINKV--LTDPK--FYNPEMOKKGAAPKPMNKDSFLYIDRSFYIL- 288
 Db 195 FPIV-----FYVSKSLKSLDDHDLFID-----TGHPPLSKALKLCFYDKTYFYLL 241
 QY 289 ----LTLTIVAYGICINLIEVTKWSQKLQVPMNNDV--EFMGNPSPFWGVSVLML 342
 Db 242 CFYFMQLALAI-----TEFNVLKIFEIPASKEEFELVAHIGKSLMISLGMICAL 294
 QY 343 FVGVNVRKFGVLTGALVTPVWVLLTGIVFPAVIFRNQASGLVAMFGTTPMLAVVGA 402
 Db 295 FAYSVIVKRLG-----VNNIILFAPLCFLS-----LFLWTFKTTLSIAVLA 336
 QY 403 IONILSKSTKALFDPSTKEMAY--IPLDQOKVKGAIDVVAARFGSGALLIQOGLLV 460
 Db 337 M--VVRGVTYALDDNNLQLLYGVP---NKIRNQ--IRIVSEFISPIGMLVMS--L 385
 QY 461 ICGSIGAMTPYLAIVLLFTIAIMLVSAIKNKLFLAQSAIK 501
 Db 386 VC--FLSSQOYVFCIIISLIATILVC--LVRSYAKAALK 421

RESULT 15

AA36963

ID AA36963 standard; Protein; 420 AA.

AA36963;

07-OCT-1999 (first entry)

Amino acid sequence of a Chlamydia trachomatis protein.

KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KW Paratrachoma; inclusion conjunctivitis; genital disease; perithenaritis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis.

OS Chlamydia trachomatis.

XX MO9928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-1B01939.

XX 04-NOV-1998; 98US-0107077.

XX 28-NOV-1997; 97ER-0015041.

XX 17-DEC-1997; 97FR-0016034.

XX (GENSET) GENSET.

XX Griffais R;

XX WPI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

XX Disclousure; Page 807-809; 1755pp; English.

CC AA36754-373949 are encoded by open reading frames (ORFs) of the genome
 CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nonendemic trachoma,

CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis, perithenaritis;
 CC Bartholinitis; pneumopathy; venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.

XX SO Sequence 420 AA;

Query Match 5.8%; Score 152; DB 20; Length 420;

Best Local Similarity 20.3%; Pred. No. 1.2e-07;

Matches 95; Conservative 74; Mismatches 170; Indels 128; Gaps 20;

QY 7 KPFGLRS--FLMFIHTHEKLVLPFMFECIFENVTVLRDTQTLIVAGPQS----- 58
 Db 15 KGSGLRALFLSLRGEKRAL-LTLGLI-----MSVACYSLAIGES 60
 QY 59 -----GAELPF-----IKFVLVPCAIIFMLIYAKLSNLKQALFYAVGTPELI--- 104
 Db 61 XFLBEIGAEKLPFAYLGSFPL---CFISCLILYNLSRKKEVSPRALFLSFICVLICNLY 117
 QY 105 -FPAFPFTVYPLRDVHPTERADRLQALIPGGLGLVALIRNWTFAFYVLAELMG-SV 162
 Db 118 LFWTL-----AIH-----KVSQTPPL-----YRL--IWGLTI 145
 QY 163 MLSLMPWGFANETKHEAKRFYALPGIGANISLASGRAIVWASK-LRASVSGVDPWG 221
 Db 146 LCVANFMGFIDQFNIDAKRHFCTF-----MAITFGDPLGARIVNQIYLG 193
 QY 222 ISRLMAMTIVSGLVIMASYWVWINKVLTDPK--FYNPEMOKKGAAPKPMNK 275
 Db 194 AELILAFIVITFIFLVHYISSLSKESEDHDLFDLTGYPSPTQ-----TLKICLK 247
 QY 276 DSFLYIDRSFYILTLTIVAYGICINLIEVTKWSQKLQVPMN--DYSEPMNPSFWT 333
 Db 248 DKYTFYLVSEFYLMQLLVFT-----EFNVLKIFDQFNAETCELTEENFTKYSWI 299
 QY 334 GVSVLMLFVGVNVRKFGVLTGALVTPVWVLLTGIVFPAVIFRNQASGLVAMFGTTP 393
 Db 300 SLGNMFPALPAYSRYVTRFG-----INNIIILFAPLCF-----PSLFCWS 339
 QY 394 LMLAVGAIQNLISKSTKALFDPSTKEMAYIPLDQOKVKGAID 440
 Db 340 IKTSVFATWGMARBEGLAVALLDNNLQLLYGIPKIRNQVMAIE 386

RESULT 16

AA74601

AA74601;

21-MAR-2000 (first entry)

Neisseria meningitidis ORF 163 protein sequence SEQ ID NO:676.

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.

XX Neisseria meningitidis.

XX MO9957280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUN-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX MPI; 2000-062150/05.
 DR N-PSDB; AA253363.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 PS
 PS Claim 2; Page 461; 1453pp; English.
 XX
 XX AA253015 to AA254536, AA254577 to AA254615, and AA774253 to AA775941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 660 AA;
 Query Match 5.5%; Score 144; DB 21; Length 660;
 Best Local Similarity 20.7%; Pred. No. 1.9e-06;
 Matches 127; Conservative 100; Mismatches 170; Indels 216; Gaps 34;
 QY 32 LMFCTITRYVYVLRDTKDTLVGAPG-----SGEAIPIKE-MLVVPICAIIFM----- 79
 Db 7 LFFVCLV-----VLVLTVPDOVMILDKAEVIFTEFSPVYLTFISIFLPII 56
 QY 80 LIYAKLSNI-----LSKQALFYAVGTFP-LIFPALFPTVYPLRDVHPTEF 125
 Db 57 LSVSLGNIRLGRDDEVEPFGFLSLMLFPAAGVGGLMFQVAPLIMHYSIDTAGTPE 116
 QY 126 ADRLQALIPGILGLVAILRMTFAFYVLAELMGVNLSTMFQFANEITKIHAKR-- 183
 Db 117 HRQOQALL-----HVFHMGVHAMS-----YGTIALALAVPGR---YKLPLALNSC 161
 QY 184 FYAL-----FGIGANI-SILASGRAIV-----WASKLRASVE--GVDPWGISLRLL 227
 Db 162 FYPLIKERKISGRFGDAIDIMALLATFFGIITLFGASQGLGAGLQEMGMIKENSFSVQL 221
 QY 228 MAMTIVSGIVLMASVWMIKNKVLTPRFRNPEMGKKGAK--PKNMKQSEF-----YL 281
 Db 222 IIAAWMSLAIVSA-----ISGVGGVYVSELNLGLAFLFFVL 261
 QY 282 DRSPYILLTLTLVAVGICI-----NLIEVTKSQSLQYPMNDYSEFMGNFSF----- 331
 Db 262 AAGFVYVLLS-----AFGDNIGNYLGNLVRLSKT-----YAKERHKWFESWTLYWAM 312
 QY 332 ---WTGVSVLIMLFVGVGNVIRKFGMLTGALVTPVWVLLTGIVFALVI FRNQASGLVAM 388
 Db 313 WCSNAPFVGLFIARISKRTIREF--VFGLIIPG-----LFGVLMF-----TV 354
 QY 369 FGTTFMLL-AVYVGAIQNLISKSTKVALFEDSKEMAVYPLDQEQV----- 433
 Db 365 FGNVIMWLDGAGVAGMLERKTSPTLTF--KFNRYPLPLISVISLVISLFFVTS 411
 QY 434 -----KGKAA-----IDVVAARFGKSGGALIQOGLLVICSGIGAM 468
 Db 412 DSGIVLVNNTSRDGLAPRWQAVMWGVLMGSAVAVLMBRSG-----LGNLSM 461

QY 469 T-----PYANVILFIATIWL-VSA-----TKLN-----KLFQAQALKEQ 503
 Db 462 TLIVSLPPLMLTMCSTLMKGLSADKKYFETRVNPTSVFWTGSKMKERLVQIMSQTOBQ 521
 QY 504 EV---AOEDSAPA 513
 Db 522 DILKFLKOTASPA 534
 RESULT 17
 ID AA774602
 ID AA774602 standard; Protein; 660 AA.
 XX
 AC AA774602;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 163 protein sequence SEQ ID NO:678.
 XX
 KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.
 OS Neisseria meningitidis.
 XX
 EN MO9957280-A2.
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 DR MPI; 2000-062150/05.
 DR N-PSDB; AA253364.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 PS
 PS Claim 2; Page 463; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AA774253 to AA775941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 660 AA;
 Query Match 5.5%; Score 144; DB 21; Length 660;

Best Local Similarity 20.7%; Pred. No. 1.9e-06;
Matches 127; Conservative 99; Mismatches 171; Indels 216; Gaps 34;
QY 32 LMFECITNTYVLRDTKDTLVAGP-----SGAEAIPIKPF-WLVVPCAIITM----- 79
DB 7 LFFVCLV-----VLVTVPDQVOMLDRAKEVITFEFSWFVLTFSIFGLPILI 56
QY 80 LIYAKLSNI-----LSKQALFYAVGTFP-LIFPALPTVYIPLRDVLPHEF 125
DB 57 LSVSSLGNIRGRDEVDPEFGFLSWLMLFAAGVGVLFFGVAPLPHMYESDITAGTPE 116
QY 126 ADRLQAIIPGGLGLVALIRNMTFAFYVLAELMGVSLMFMGFANEITKIHAKR-- 183
DB 117 HRQOQALL-----HTVHMVGVAHSV-----YGTIALALAYRGR--YKLPLALRSC 161
QY 184 FYAL-----FGIGANI-SLASGRAIV-----WASKLRASVE--GVDPWGISLRLL 227
DB 162 FYPLKEKISGRFGDAIDIMALLATFFGIITLFGASQGLGAGIQEIGWIAENSFSVQVL 221
QY 228 MAMTIVSGLVLMASYWINKVLTDPFRYPNEEMQKKGAK--PKNMKDSFL-----YL 281
DB 222 IIAVMSLAVVSA-----ISGVGKGVVLSLNLGLAFLLLPFLV 261
QY 282 DRSPYILLTLTLVAVGICI-----NLIEVTKSOLKIQYPMNDYSEFMGNFSF----- 331
DB 262 AAGPTVYLLS-----AFGDNIGNYLGNIYRLSFKT-----YAYERHKKWPFESWTLYWAW 312
QY 332 ---WTGVSVLIMLFVGVGNVIRKPGMLTGALVTPVMVLLTGIVFPALVIFRNQASGLVAM 388
DB 313 WCSWAPFVGGLFIARISKRTIREF--VFGVLLIPG--LFGVLMF-----TV 354
QY 369 FGTPPLML-AVVVGAIONILSKSTKVALFDSTKEMAYIPLDOEQV----- 433
DB 365 FGNIAIMNDGVAGVLEKMTISSPETLLF--KFNYPILPELISIVSLVISLFFVTSA 411
QY 434 -----KGKAA-----IDVVAARFGKSGGALIIOGGLLVICSGIGAM 468
DB 412 DSGIYVNNITSRDGLSAPRWQAVMWGVMSAVAVLMSRSG-----LGNLQSM 461
QY 469 T-----PYLAVILLFIAIWL-VSA-----TKLN-----KLFASALKEQ 503
DB 462 TLIVSLPFPALMLMPCSLMKGLSADKKYFETRVNPTSVFTGCKMKERLVQIMSOQEO 521
QY 504 EV---AOEDSAPA 513
DB 522 DILKFLKHTASPA 534

RESULT 18
AAU91074
ID AAU91074 standard; Protein; 675 AA.
XX AAU91074;
XX
XX 05-JUN-2002 (first entry)
XX
XX Neisseria cell surface polypeptide #12.
XX
XX Cell surface protein; antibacterial; antimicrobial.
XX
XX Neisseria meningitidis.
XX
XX WO200216612-A2.
XX
XX 28-FEB-2002.
XX
XX 21-AUG-2001; 2001WO-GB03759.
XX
XX 24-AUG-2000; 2000GB-0020952.
XX
XX (MTCR-) MICROSCIENCE LTD.
XX
XX Lane JD, Hughes MJG, Santangelo JD;
PI

XX WPI: 2002-280941/32.
DR N-PSDB; ABE54090.
XX
XX Novel peptide encoded by Neisseria meningitidis, useful for manufacture
PT of medicament for treatment or prevention of condition associated with
PT infection by Neisseria or Gram-negative bacteria -
XX
XX Claim 5; Page 60-62; 79pp; English.
XX
XX The invention relates to polypeptides located on the cell surface of
CC Neisseria meningitidis, and the polynucleotides encoding them. The
CC sequences of the invention are useful for therapeutic or diagnostic use,
CC in the manufacture of a medicament for use in treatment or prevention of
CC a condition associated with infection by Neisseria or Gram-negative
CC bacteria. The sequences are also useful for screening potential
CC antimicrobial drugs or for detection of virulence. Sequences
CC AAU91063-AAU91079 represent Neisseria meningitidis polypeptides of the
CC invention.
SQ Sequence 675 AA;
Query Match 5.5%; Score 144; DB 23; Length 675;
Best Local Similarity 20.7%; Pred. No. 1.9e-06;
Matches 127; Conservative 100; Mismatches 170; Indels 216; Gaps 34;
QY 32 LMFECITNTYVLRDTKDTLVAGP-----SGAEAIPIKPF-WLVVPCAIITM----- 79
DB 22 LFFVCLV-----VLVTVPDQVOMLDRAKEVITFEFSWFVLTFSIFGLPILI 71
QY 80 LIYAKLSNI-----LSKQALFYAVGTFP-LIFPALPTVYIPLRDVLPHEF 125
DB 72 LSVSSLGNIRGRDEVDPEFGFLSWLMLFAAGVGVLFFGVAPLPHMYESDITAGTPE 131
QY 126 ADRLQAIIPGGLGLVALIRNMTFAFYVLAELMGVSLMFMGFANEITKIHAKR-- 183
DB 132 HRQOQALL-----HTVHMVGVAHSV-----YGTIALALAYRGR--YKLPLALRSC 176
QY 184 FYAL-----FGIGANI-SLASGRAIV-----WASKLRASVE--GVDPWGISLRLL 227
DB 177 FYPLKEKISGRFGDAIDIMALLATFFGIITLFGASQGLGAGIQEIGWIAENSFSVQVL 236
QY 228 MAMTIVSGLVLMASYWINKVLTDPFRYPNEEMQKKGAK--PKNMKDSFL-----YL 281
DB 237 IIAVMSLAVVSA-----ISGVGKGVVLSLNLGLAFLLLPFLV 276
QY 282 DRSPYILLTLTLVAVGICI-----NLIEVTKSOLKIQYPMNDYSEFMGNFSF----- 331
DB 277 AAGPTVYLLS-----AFGDNIGNYLGNIYRLSFKT-----YAYERHKKWPFESWTLYWAW 327
QY 332 ---WTGVSVLIMLFVGVGNVIRKPGMLTGALVTPVMVLLTGIVFPALVIFRNQASGLVAM 388
DB 328 WCSWAPFVGGLFIARISKRTIREF--VFGVLLIPG--LFGVLMF-----TV 369
QY 369 FGTPPLML-AVVVGAIONILSKSTKVALFDSTKEMAYIPLDOEQV----- 433
DB 370 FGNIAIMNDGVAGVLEKMTISSPETLLF--KFNYPILPELISIVSLVISLFFVTSA 426
QY 434 -----KGKAA-----IDVVAARFGKSGGALIIOGGLLVICSGIGAM 468
DB 427 DSGIYVNNITSRDGLSAPRWQAVMWGVMSAVAVLMSRSG-----LGNLQSM 476
QY 469 T-----PYLAVILLFIAIWL-VSA-----TKLN-----KLFASALKEQ 503
DB 477 TLIVSLPFPALMLMPCSLMKGLSADKKYFETRVNPTSVFTGCKMKERLVQIMSOQEO 536
QY 504 EV---AOEDSAPA 513
DB 537 DILKFLKHTASPA 549

RESULT 19
AAU74600

ID AAY74600 standard; Protein; 660 AA.
 AC AAY74600;
 DT 21-MAR-2000 (first entry)
 DE Neisseria gonorrhoeae ORF 163 protein sequence SEQ ID NO:674.
 KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KM antibacterial; gene therapy.
 OS Neisseria gonorrhoeae.
 PN MO9957280-A2.
 PD 11-NOV-1999.
 PF 30-APR-1999; 99WO-US09346.
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 DR WPI: 2000-062150/05.
 DR N-PSDB; AA253362.
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 PS Claim 2; Page 460; 1453pp; English.
 CC AA253015 to AA254536, AA254537 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 SQ Sequence 660 AA;
 Query Match 5.1%; Score 134; DB 21; Length 660;
 Best Local Similarity 20.3%; Pred. No. 2.4e-05;
 Matches 125; Conservative 100; Mismatches 169; Indels 222; Gaps 35;
 QY 32 IMFPCITFNYVLRDKTLIVAGPQ-----SGAEAPIFIRF-MLVVPCALIFM----- 79
 DB 7 LFFVCLV-----VLVLTVPDQVQMWLDRAKEVIFTEFSWYVLTFSIFGLLI 56
 QY 80 LIYAKLSN-----LSKQALFYAVGTFP-LIFPALFPTVIVPLADVLAHPTTF 125
 DB 57 LSVSGIGNIRLGRDDEVPFGFLSWLAMPFAAGMGVGLMFVVAEPLMHYFSDIVGABE 116
 QY 126 ADRLQALIFPGLLGVAVILRNWTFAPAFVYLAELMGSVMLFMFGPANEITIHAKR-- 183

DB 117 HRQQQALL-----HTVFHGVHAWSV-----YGTIALALAYFGFR---YTLPLALNSC 161
 QY 184 FYAL-----FGIGANI-SLASGRAIV-----WASKLASVSE-----GVDPMGISLR 225
 DB 162 FYPILKEKISGRFGAIDIMALLATFGIITTLGFGASQLGAGLOEMGMIAENSFGVYL 221
 QY 226 LL---MAMTIYSGVLMSYMMINKGNVLTDPFRFYPEEMOGKKGAKPRMKNKDSFL--- 279
 DB 222 IIAVMSLAIVSAI-----SGVGKGVK-VISEINIGLAFLLF 258
 QY 280 -YLDSPYILLTLTLVAVGICI-----NLEVTWKSQKQYPNMNDYSEFMGNSF-- 331
 DB 259 FYLAADPTVYLLS---AFGDNIGNVLDNVLVRLSKT-----YAYEREHKPFESMTVLY 309
 QY 332 -----WTGVSVVLIMLVGVGNVIRKFGMLTGALVTPVMWLTGIVFPALVIFRQASGL 385
 DB 310 WAMMCSWAPFVGLFIRISKRTIRF--VFGVLLIPG---LFGVLMF----- 352
 QY 386 VAMGCTPLML-AVVVGAIONILSKTKYALFDSFKEMAYIPLDEOKV----- 433
 DB 353 -TVFGNTAIWLDGVAGMLERKMTSSPETLFP--KFTNYLPPLTLSTIVSLIVSLFV 408
 QY 434 -----KGKAA-----IDVVARFGKSGGALIQGGLVYCGSI 465
 DB 409 TSADSGIVVANNITSRDGLSAPRMQAVMMGVMSAVVILMRSG-----LGNTL 458
 QY 466 GANT-----PYLAVILFIIAIWL-VSA-----TKLN-----KLFLAQSAL 500
 DB 459 QSMTLIVSLPFLMLIMCFSLMKGLSADKKYFETRVNPTSVFMTGKWKERTLVIRMSOT 518
 QY 501 KROEV---AOEDSAPA 513
 DB 519 QEDILKELKHTASPA 534
 RESULT 20
 ABP78636
 ID ABP78636 standard; Protein; 675 AA.
 AC ABP78636;
 DT 07-MAR-2003 (first entry)
 DE N. gonorrhoeae amino acid sequence SEQ ID 3802.
 KM Antibacterial; infection; vaccine; gene therapy.
 OS Neisseria gonorrhoeae.
 PN WO200279243-A2.
 PD 10-OCT-2002.
 PF 12-FEB-2002; 2002WO-1B02069.
 PR 12-FEB-2001; 2001GB-0003424.
 PA (CHIR-) CHIRON SPA.
 PI Fontana MR, Pizsa M, Masignani V, Monaci E;
 DR WPI: 2003-058415/05.
 DR N-PSDB; AB239606.
 PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection -
 PS Disclosure; Page 463; 815pp; English.
 CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition

completing the protein, nucleic acid or antibody is useful for the manufacture of a medicament for treating or preventing N. gonorrhoeae infection, this may be in the form of a vaccine or gene therapy. Sequences given in records ABP76736-ABP81046 represent nucleic acid molecules of the invention.

Sequence 675 AA;

Query Match 5.1%; Score 134; DB 24; Length 675;

Best Local Similarity 20.3%; Pred. No. 2.5e-05; Indels 222; Gaps 35;

Matches 125; Conservative 100; Mismatches 169;

32 IMFPCITNYTVLRTKDTLVGAGC-----SGAEALPIFIK-WLVVPCATIFM----- 79
 22 LFFVGVLV-----VLVITVPDQVMMDRAKEVIFTEFSFYVLTSEIFIGFILLI 71
 80 LIYAKLSNI-----LSKQALFYAVGTPE-LIFPALFTVIVPLDVLHPTPEF 125
 72 LSVSGLGNIRLGRDEDEVEFGLSWLAMLFAAGMGVGLMFVABPLMHFSDITVGAPE 131
 126 ADRLOAIPFGILGLVALRWNTFAAFYVLAELMSVMSLMFNGFANEITKHEAKR-- 183
 132 HRQOQALD-----HTVFHMGVHMSV---YGTIALALAYFGFR---YKLPLALMSC 176
 184 FYAL-----FGIGANI-SLLASGRAIV-----WASKLRASVE-----GVDWPGISLR 225
 177 FYPLIKKISGRFGDAIDIMALLATFFGLITTLTGSGASQLGSLGLOEMGMIENSGVYL 236
 226 LL---MAMTIVSGVLMSYWMINKNVLTDPRFYNPEEMQKKGAKPRMNKDSFL--- 279
 237 IIAAVMSLAIVASAI-----SGVGKGVK-VLSELNIGLAFLLLF 273
 280 YLDRSPITLLTLTLVAVGICI-----NLIEVTKKSQKIQYPMNDYSEPMGNEFR-- 331
 274 FVLADPVPVLLS---AFGDNIGNVLTGLVRLSLKT-----YAYERHKKPFESWTLY 324
 332 -----WTVGVSVLMLFVGVNVIRKFGMLTGCALTVPVWVLTGIVFPALVIFRQASGL 385
 325 WAMMCSNAPVGLPIARISKRTREF--VGVLLIIPG---LFGVLM----- 367
 386 VAMFGTTPMLT-AVVVGAIONILSKSTYALPDSTKENAYIPLDQEQRY----- 433
 368 -TVGNTAIMLNDGVAAGMLEMTSSPELTLLF---KFPNYPILPELTISIVSLVLSLFPV 423
 434 -----KGKAA-----IDVNAARCKSGGALIQGLLVICSI 465
 424 TSADSGIYVNNITSRDGSLAPRQAVMVGVLMSAVAVLLMRSG-----LGNTL 473
 466 GAWT-----PYLAVILFLIAIWL-VSA-----TKLN-----KLFLAQSAL 500
 474 QSMTLIVSLPFLALMLIMCPSLMKGLSDKKYFETRVNVPVFWTGKMKERLVIRMSQT 533
 501 KEQEV---AQEDSAPA 513
 534 QEQDILKFLKHTASPA 549

RESULT 21
 ID AU34551 standard; Protein; 463 AA.
 AC AU34551;
 DE 14-FEB-2002 (first entry)
 XX E. coli cellular proliferation protein #132.
 KM Antisense; prokaryotic cellular proliferation protein;
 XX antibiotic; antibacterial; drug design.
 OS Escherichia coli.
 XX
 XX WO200170955-A2.

XX 27-SEP-2001.
 PD 21-MAR-2001; 2001WO-US09180.
 XX 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX (ELITRA PHARM INC.
 PA Haeelbeck R, Ohlsen KU, Zyckind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 PL WPI; 2001-611495/70.
 DR N-PSDB; AAS52410.
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PS Example 3; Seq ID No 10144; sllp; English.
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are:
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 463 AA;
 SQ
 Query Match 5.0%; Score 132.5; DB 22; Length 463;
 Best Local Similarity 22.3%; Pred. No. 2.1e-05;
 Matches 108; Conservative 69; Mismatches 174; Indels 133; Gaps 28;

69 WLVVPCATIFMLIVAKLSNI-LSKQALFYAVGTPELIFPALFTVIVPLDVLHPTPEAD 127
 39 WISVPCLLAFVCVWMLRSAAVAVNLPKGFNFTTQQLFMLTRALPVSGLALRV--PYSG-- 94
 128 RLQALIPRGLGLVAL--RWTFPAFYVLAELMSVMSLMFNGFANEITKHEAKRFY 185
 95 -----WVPIFGRRMT--AFST-----GIILIPCVLWGF-----VQDSTPY 130
 186 ALFGIGANISLLASGRAIVWASKLRASVS-----EGVDWPGISRLIMAM 230
 131 SVFII---ISLCCFAGANPSSM-ANISFFPKQKQGALGLNGIGMNCVSMQVLAP 186
 231 TIVSGVLWMSYWMINKNVLTDPRFYNPEEMQKKGAKPRMNKDSFLYDRSPYLL- 289
 187 LVWS-LSIFAVF-----GSGGVKQPDGTE--LYLANASWIVWP 221
 290 -LTLIVAYGICINLVTKKSQKIQYPMND-----YSEFMGNF-SFTYGV--- 335
 222 FLAIFTIAWFGMNDL-ATSKASIKQPLVKRGLWIMSLVLTATGSGFISAGPAML 280
 336 -----VSVLIMLVFG---GNVIRKFGW-----LTGALVTPV---MVLITGIVFPALV 376

Db 281 SKQGPVQILQVAFGPFIGALARSAGALSDRLGGRVTLVNTLMAIFSGLLFTLP 340
 QY 377 IFRNQASLVAMFGTPTMLAVV-----GAIQNIISKSTKTKLFDSTKEMAYIPLDQE 430
 Db 341 T-DGGGSGFMAFFA---VFLLFLTAGLGGSTFQMSIVIRKULTMDRVKAEQ---GSD 392
 QY 431 QKVQKKAIDVVA-----RFGSGGALLIOGL---LVICGS-IGAMTPYLAVILLFPIA 481
 Db 393 ERARKERATDTAALGFIISAIGAGFTIPKAFGSSLLATGSPVGAMKVFILFYACVVI 452
 QY 482 IWL 485
 453 TWAV 456

RESULT 22
 AAU34568
 ID AAU34568 standard; Protein; 462 AA.

XX AC AAU34568;
 XX DT 14-FEB-2002 (first entry)
 XX DE E. coli cellular proliferation protein #149.
 XX KM Antisense; prokaryotic cellular proliferation protein;
 XX KM antibiotic; antibacterial; drug design.
 XX OS Escherichia coli.
 XX PN W0200170955-A2.
 XX PD 27-SEP-2001.
 XX PF 21-MAR-2001; 2001WO-US09180.
 XX PR 21-MAR-2000; 2000US-191078P.
 XX PR 23-MAY-2000; 2000US-206848P.
 XX PR 26-MAY-2000; 2000US-207727P.
 XX PR 23-OCT-2000; 2000US-242578P.
 XX PR 27-NOV-2000; 2000US-253625P.
 XX PR 22-DEC-2000; 2000US-257931P.
 XX PR 16-FEB-2001; 2001US-263308P.
 XX PA (BLIT-) BLITRA PHARM INC.
 XX PI Haeelbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 XX PI Yamamoto RT, Xu HH;
 XX DR WPI; 2001-611495/70.
 XX DR N-PSDB; AAS52427.
 XX PT New polynucleotides for the identification and development of
 XX PT antibiotics, comprise sequences of antisense nucleic acids -
 XX PS Example 3, Seq ID No 10161, 511pp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 462 AA;
 Query Match 4.9%; Score 129; DB 22; Length 462;
 Best local similarity 22.3%; Pred. No. 5,1e-05;
 Matches 108; Conservative 71; Mismatches 171; Indels 134; Gaps 26;

QY 69 WLVPICAIIFMLIYAKLSNI---LSKQALFYAVGTPFLIFALPPTV-----IY 114
 Db 37 WISVSCILLAFQVMMLFSATVNLKIGFNTTQTLT---TALPSVSGALLRPYSRMV 94
 QY 115 PLRDVLHTEFADRLQAILPPGLIGLVAILLNMTFAAPFYVLAELWGSVMSLMFWG--FA 172
 Db 95 PIFGRRWTFVSTAI-LIIPCVMLGIAVQNPNTFFGIFIVIALCG-----FAGANFA 146
 QY 173 NEITKIH---EAKRFYALRGIGANISLSGRAIWAASKURASVSEGVDPWGISLRLLM 228
 Db 147 SSMGNISFFPEPKAKQSGAL-GINGGL-----GNLGSVMQULVAPLVFVPEFA 193
 QY 229 AMTI-----VSGVLMASVW-----WINKVLTDPFRFYNPEMOKKGA 268
 Db 194 FLGVNGVPQADGSVMSLANNAWIVPPLATATITANAGANDI-----A 236
 QY 269 KPRKMKDSFLYLDSPYILLTLVIA-YGICINLIEVTKWSQKLQYPMN-----DY 322
 Db 237 SSRASIQDLPLQRL-HLWLLSLVLAFTGSGFIG-PSAGPAMAKQFPDVNILLRLAF 294
 QY 323 SEFWGNFSFWTGVSVLIMLFVGGNVIRKGMULTGALVTPVMULT-----TG-- 369
 Db 295 GPFIG-----AIAHSVGAISDKFGKGVTLINIFMAIFALLFLLPGTSG 343
 QY 370 --IYFPALVIFRNQASGVAMFGTPTMLAVVVAIGAIQNIISKSTKY--ALFDSKEMAYI 425
 Db 344 NFIAFYAVPMGLFTIAGIGS--GSTFQMAIVIFQI-----TIYKVMKGGSDQAH- 393
 QY 426 PLDQQRVKKKAIDVVAARFGKSGALLIOGL---LVICGS-IGAMTPYLAVILLFPIA 481
 Db 394 ---KEAVTEPFAALGFIISA-IGAVGPFITPQAFGMSLMNMGSPVGAMKVFILFYACVIL 449
 QY 482 IWL 485
 450 TWLV 453

RESULT 23
 ABJ19074
 ID ABJ19074 standard; Protein; 435 AA.
 XX AC ABJ19074;
 XX DT 06-MAR-2003 (first entry)
 XX DE Pathogen specific antigen related staphylococcal protein SEQ ID No 351.
 XX KM Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
 XX KM hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
 XX KM auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
 XX KM autoimmune disease; HIV; hepatitis.
 XX OS Staphylococcus sp.
 XX PN W0200259148-A2.
 XX PD 01-AUG-2002.
 XX PF 21-JAN-2002; 2002WO-EP00546.
 XX PR 26-JAN-2001; 2001AT-0000130.

(CIST-) CISTEM BIOTECHNOLOGIES GMBH.

PA Meinke A, Nagy E, Von Ahnen U, Klade C, Henics T, Zauner W,
PI Minn DB, Vytvytska O, Etz H, Dryla A, Weichart T, Hafner M,
PI Tempelmeier B;

DR WPI, 2003-075410/07.

PT Identifying, isolating and producing hyperimmune serum-reactive
PT antigens from a pathogen, for preparing vaccine or medicament for
PT treating or preventing e.g. staphylococcal infections, comprises
PT providing antibody preparation

PS Example 7; Page 214; 252pp; English.

CC The invention relates to a novel method for identifying, isolating and
CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,
CC allergen, a tissue or host prone to auto-immunity, where the antigens
CC are used in a vaccine, comprises providing antibody preparation from a
CC plasma pool of a type of animal, or individual sera with antibodies
CC against the specific pathogen, tumour, allergen, tissue or host prone to
CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of
CC the 62 sequences of 53-2261 amino acids fully defined in the
CC specification, or their hyperimmune fragments are useful for the
CC manufacture of a pharmaceutical preparation, particularly a vaccine
CC against staphylococcal infections or colonisation against S. aureus or S.
CC epidermidis. The preparation of antibodies is useful for the manufacture
CC of a medicament for treating or preventing staphylococcal infections or
CC colonisation against S. aureus or S. epidermidis. The antibody
CC preparations may also be used for diagnostic and imaging purposes. Other
CC conditions that can be treated include cancer, autoimmune diseases or
CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or
CC protozoan pathogens. This sequence represents a staphylococcal protein
CC relating to the method for identifying and producing pathogen specific
CC antigens of the invention.

XX Sequence 435 AA;

Query Match 4.8%; Score 125.5; DB 24; Length 435;

Best Local Similarity 19.9%; Pred. No. 0.00011; Mismatches 162; Indels 143; Gaps 19;

Matches 96; Conservative 82; Mismatches 162; Indels 143; Gaps 19;

9 FGKLSFMPITHTLTKKVLPMFPCITFNYTVLRKOTLIYAGPQSGAEI-----63
DB 17 FGAGLILPMPMIGHAGQMMWIGMGPALT---GILLPITYIVAFDEGVESGNRIH 73
QY 64 PFKEWLVVPCAIIFMLIYA-----KLSNISKQALFYAVGT-----PFLIPFAL 108
DB 74 PWFGRIFAV---VIYMSIGAFYGIIPRAANVA-----YEIGRRHILPVHNGWTLIIFAI 124
QY 109 PFTVYVPPADVLHPEFADRLCAIIPGGLGVALIKRMTFAFV-----153
DB 125 FFALVYWIS--LNPSTIVDNIGKLLTPLLLMVALL---SLAVINPESASAPROKXIT 179
QY 154 -----VLAEIMGVSV-----MLSLMFGFANEITKIHAKRF 184
DB 180 HPFISGLEGYFTTMVLVAALAEVYIVNGYKXKGLTDMKIKLYCFSGLIALLIGMIV 239
QY 185 YALFEGIANIS--LIASGRAIWAASKLRASVEGVDPWGISIRLLAMTIVSGVLWASY 242
DB 240 FALAVAGASTAGNFKDGTDLITVNSLRFGSFGMLVFGMTV--ILACTTTCIGLV--NACA 297
QY 243 WMINNVNLTDRFRVPEEMOKKKAKPRMNMK-----DSFYLDKSPYIIL--LT 291
DB 298 TFFKGVH-----PKFSKIPALIFSIIGFLFTLIGEMILIKIAPV 337
QY 292 LLIYAVGICINLIETWKSQKLQYPMNDYSEFGNNGSEFMTGVSVLIMFVGNAVIRK 351
DB 338 LITLIPSTIALVLSIF-----ANMSTFFPSNAYRIATYITLITSL--QIINS 385
QY 352 FGMLTGALVTPVMVLLTGIVFPALVIFRNOASGLVAMEGTTPLMLAVVGAIONILSKST 411
DB 386 FNLHGVILKSFMMLEPLADIDLAMLV-----PFMLFALIGFIIDVFIRP 430

QY 412 KYA 414
DB 431 KQA 433

RESULT 24

ID AAY58582 standard; Protein: 713 AA.

AC AAY58582;

DT 10-APR-2000 (first entry)

DE Sorangium cellulosum protein Orf 3.

KW Orf 3; Na/H antiporter; epothilone biosynthesis; export;
KM taxol substitute; anticancer.

OS Sorangium cellulosum.

FN W09966028-A2.

PD 23-DEC-1999.

PF 16-JUN-1999; 99WO-EP04171.

PR 18-JUN-1998; 98US-0099504.

PR 24-SEP-1998; 98US-0101631.

PR 05-FEB-1999; 99US-0118906.

PA (NOVS) NOVARTIS AG.

PI (NOVS) NOVARTIS-ERRINDUNGEN VERM GES MBH.

PI Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;

DR WPI; 2000-097741/08.

XX N-PSDB; AAY58587.

PT New isolated epothilone synthase genes, used for the recombinant

PT production of epothilone for use in cancer therapy

PS Claim 12; Page 162-164; 174pp; English.

CC This sequence represents a protein, Orf 3, encoded by open
CC reading frame 3 (ORF3) of a Sorangium cellulosum 68.75 kb contig
CC which covers the epothilone biosynthesis genes. This sequence has
CC homology to Na/H antiporters of various bacteria, and may take
CC part in the export of epothilones from the producer cell. The
CC invention relates to nucleic acid sequences encoding epothilone
CC biosynthetic enzymes from Sorangium cellulosum. Epothilones A and B are
CC 16-membered macrocyclic polyketides with an acylcysteine-derived
CC starter unit; polyketides being synthesised from two-carbon building
CC blocks, the beta-carbon of which always carries a keto group. Each round
CC of the two-carbon addition is carried out by a complex of enzymes known as
CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC and EPOS E (AAY58575-58578) are involved in polyketide backbone
CC formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC involved in transport. Epothilones mimic the biological activity of
CC taxol, and may be substituted for taxol in cancer chemotherapeutic
CC compositions. Epothilones exhibit a much lower drop in potency against a
CC multiply drug-resistant cell line compared with taxol, and are
CC considerably less efficiently exported from such cells by the multidrug
CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC epothilones as anticancer agents, they are problematical to produce on a
CC large scale. Epothilones are too complex for industrial scale chemical
CC synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC poor yields of epothilones. The nucleic acids of the invention may be
CC used for the recombinant production of epothilones in a heterologous host
CC that is more amenable to fermentation. Sequences AAY58580,

PT antigens from a pathogen, for preparing vaccine or medicament for
PT treating or preventing e.g. staphylococcal infections, comprises
PT providing antibody preparation -
XX
XX
PS Example 7, Page 214, 252pp; English.
CC The invention relates to a novel method for identifying, isolating and
CC producing hyperimmune serum-reactive antigens from a pathogen, tumor,
CC allergen, a tissue or host prone to auto-immunity, where the antigens
CC are used in a vaccine, comprises providing antibody preparation from a
CC plasma pool of a type of animal, or individual sera with antibodies
CC against the specific pathogen, tumor, allergen, tissue or host prone to
CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of
CC the 62 sequences of 53-2261 amino acids fully defined in the
CC specification, or their hyperimmune fragments are useful for the
CC manufacture of a pharmaceutical preparation, particularly a vaccine
CC against staphylococcal infections or colonisation against *S. aureus* or *S.*
CC epidermidis. The preparation of antibodies is useful for the manufacture
CC of a medicament for treating or preventing staphylococcal infections or
CC colonisation against *S. aureus* or *S. epidermidis*. The antibody
CC preparations may also be used for diagnostic and imaging purposes. Other
CC conditions that can be treated include cancer, autoimmune diseases or
CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or
CC protozoan pathogens. This sequence represents a staphylococcal protein
CC relating to the method for identifying and producing pathogen specific
CC antigens of the invention.
XX
SQ Sequence 472 AA;
Query Match 4.5%; Score 119.5; DB 24; Length 472;
Best Local Similarity 19.4%; Pred. No. 0.00059;
Matches 58; Conservative 61; Mismatches 111; Indels 69; Gaps 12;
QY 151 AFVYLAELWGSVMSLWFGFANEITKHEAKRFPALFG--IGANISLISAGRAIWASK 208
DB 175 AFLVFTTEQGNLITAAAF-----LTAAGNPLAQNILASSTNVHITWNNW 219
QY 209 LRASVSEGVDPWGSIRLIMAMTIVSGVLMAVYWKINKVLTLPFRFNPEMOKGKKGA 268
DB 220 FLALVLP-----LVSLIVPFTI---KYIYPVKEPT--NAKSMANENLAT 262
QY 269 KPKANMKOSFLYLDSPYLLLTLLVAVGICINLIEVTWKSQKLQYPMNDYSEPMGN 328
DB 263 MGKIALAEKMT---GIFVVALTITVGSFTHIDTLTAFTALALLLTGVLTMODITNE 319
QY 329 FSEFTGVV--SVLIMLFGVGNVIRKFG--WLTGALVTP-----VMVLITGIVFPAL 375
DB 320 TGAANNLWVFSVLVLM--ADQLNKLGFIPWLSKSIATSLGSLGMPVILVITLIFEFYSH 376
QY 376 VIFPNQASGLVAMF-----GTPPLMAVVVGAIONILSKSTKA-----LFPSS 418
DB 377 YLFASSTAHISAMYAALLGVALAAGAPPLFSALMLGFGGNLLASTTHYSSGPAPILFSS 435
RESULT 28
ID ABP40810 standard; Procein; 800 AA.
XX
XX ABP40810;
XX
XX 24-JUL-2002 (first entry)
XX
XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5655.
XX
XX Staphylococcus epidermidis, open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy.
XX
XX Staphylococcus epidermidis.
XX
XX US6380370-B1.
XX
XX 30-APR-2002.

PF 13-AUG-1998; 98US-0134001.
XX
XX 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
XX WPI; 2002-381255/41.
DR N-PSDB; AEN93355.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX Disclosure; SEQ ID 5655; 267pp; English.
XX
PS AEN90538 to AEN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP5124 to ABP37960. The *S. epidermidis* sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly *S. epidermidis* infections. The sequences can be used to
CC screen for compounds able to interfere with the *S. epidermidis* life
CC cycle or inhibit *S. epidermidis* infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 800 AA;
Query Match 4.5%; Score 119.5; DB 23; Length 800;
Best Local Similarity 20.3%; Pred. No. 0.0013;
Matches 108; Conservative 86; Mismatches 148; Indels 191; Gaps 30;
QY 2 TKTEKEPFGKLSRLW-----PI-----HTHEKKVLPFIMF-----F 35
DB 222 TKSAQFPF-----HIMLPKMAAPTPVSAYHSAHTMKA-GIFLFRPTPLGLSNVIY 275
QY 36 CTFP-----NYTVLR-DYKDTL-----LVGAPSGAE-AIPFKFW 69
DB 276 TVTFVGLITMDFGLTLRQDLKGLIAYSTISQLGIMTMWVGAGGVAQHTSDELSKEY 335
QY 70 LVVFCALIFMLITAKLSNLSKOLFPYAV----- 98
DB 336 ILVLFAGLFLH---NMHAFKCALFMGVGIIHESGTRDIRLLNGKRYEPPKHIVML 391
QY 99 -----GTPFLIFPALPPTVYPRDVLPHTPEFADRLQALIPGL--LGLVALIRNMTFA 150
DB 392 AALSMAGVFPNLGLSLSEMPF---DSLTKRANLDOGFVITPVTISIGVASLITFTYA 447
QY 151 AFVYLAELWGSVMSLWFGFAN---EITKHEAKRP-----VALFGIGA 192
DB 448 LY-----MIKETFGVNYNIEKFKRKQIHBPMLFSLPAVITMLLIPVIFPVNFG 497
QY 193 NISLASGRAIWAASKLRASYS---EGVDPK-GISLALLMA-KTIVSGVLVMASTW--- 244
DB 498 NFVILPAPTRSV---SGIAGEVDAPVPHISQHWGNLPLISLIVIIIGLIALLVNMRKV 554
QY 245 -----INKNVLTDP--PRFYNPEMOKGKAGKPKANMKOSFLYLDSPYLLLTLL-----V 294
DB 555 THQIITKASITDGYRKIRYREFELYS-A-RGIRALMNANKN-----YIMTITLFIPIAI 605
QY 295 IAVGICINLIEVTWKSQKLQYPMNDYSEPMGNFSEFTGVSVLIMLFGVGNVIRKFGW 354
DB 606 VVYG-----YLVGFPHVHQL--HISFGPLEVILSV-VLLIIGISLIFIROR 650
QY 355 LTGALVTPWVYLLTGLVFPALVITFRNOASGIVAMFGTTPPLMALVVGAIQNIL 407
DB 651 LT-----NVVLNMGIMGFAYTV-----FIAMKAPDLALTOLVETITITL 690
RESULT 29

XX 14-FEB-2002 (first entry)
 DT Staphylococcus aureus cellular proliferation protein #458.
 XX
 XX
 DB Staphylococcus aureus cellular proliferation protein;
 XX
 XX Antisense; prokaryotic cellular proliferation protein;
 XX
 XX antibiotic; antibacterial; drug design.
 OS Staphylococcus aureus.
 XX
 XX WO200170955-A2.
 XX
 XX 27-SEP-2001.
 XX
 XX 21-MAR-2001; 2001WO-US09180.
 XX
 XX 21-MAR-2000; 2000US-191078P.
 XX
 XX 23-MAY-2000; 2000US-206848P.
 XX
 XX 26-MAY-2000; 2000US-207727P.
 XX
 XX 23-OCT-2000; 2000US-242578P.
 XX
 XX 27-NOV-2000; 2000US-253625P.
 XX
 XX 22-DEC-2000; 2000US-257931P.
 XX
 XX 16-FEB-2001; 2001US-269308P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 PA
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GI;
 P1 Yamamoto RT, Xu HH;
 P1
 XX WPI; 2001-611495/70.
 XX
 XX N-PSDB; AAS52041.
 DR
 XX
 XX New polynucleotides for the identification and development of
 FT antibiotics, comprise sequences of antisense nucleic acids -
 PT
 XX
 XX Example 3; Seq ID No 5678; 511pp; English.
 PS
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 XX Sequence 430 AA:
 SQ
 Query Match 4.5%; Score 117.5; DB 22; Length 430;
 Best Local Similarity 19.3%; Pred. No. 0.00087;
 Matches 93; Conservative 86; Mismatches 159; Indels 143; Gaps 19;
 QY 9 FGKLSPLMPTHTLKVLPWLMFPCITFYTYVLRDTKTLIVGAPSGAEAI----- 63
 DB 15 FGAGNLIRPPMIGTAGQNMWIGMGFALT---GILLPTITVIVAFVDEGVESYGNH 71
 QY 64 PFIRKWLVPICAIIMLYA-----KLSNYSKQLPFAVGR-----PFIIFPAL 108
 DB 72 PWFGRIFAV---VITMSIGAFYGIIPRANVA-----YEIGRHLIPVHNQMTLIIFAV 122
 QY 109 PFTVYIPRLDVLPHTPDRLOAILPQGLIGVAILRMNTFAAF----- 153

DB 123 FFAIVWIS--LNSPKIVDNLGLTLPILLMALL---SIAVIFNPESALSAPKDYIT 177
 QY 154 -----VLAELMGSYML-----SLMWRGPANETTKIHEAKRF 184
 DB 178 HPFISGLEGYFTMDLVAAAFSVIVVGYKFGLTDRTKILKYVCFSGFIAAILLGMVY 237
 QY 185 YALGIGANIS--LLASGRAIVWASKLRASVSEGVDPGISLRLMAMTIVSGVLMASY 242
 DB 238 FALAVYGASTAPGNFKOSTDILTTNSLVRGSGFQNLVFGMTV-ILACITTCIGLV-NACA 295
 QY 243 WINKQVLTDRPFYNPPEMQKKGAKPKMMK-----DSFLYIDRSPIYL--LT 291
 DB 296 TPTKKHV-----PKFSYKIFALVFISIIIGFLLGLEMIILKIAVP 335
 QY 292 LLVIAAGICINLIEVTWKSQKLQYPMNDYSEFMGNSEFWGVSVIMLFVGNVIRK 351
 DB 336 LLTLIYPSIALVLISF-----ANMSTRFSWAVYRLATVTLIIISIL-QILNS 383
 QY 352 FGMITGALVTPVWVLTGIVFPAIVIFRNQASGLVAMFGTPLMLAVVGAIONILSKST 411
 DB 384 FNLHGVIILKMFPMPLADIDLAMLV-----PFIILATIGFIIDVIRRP 428
 QY 412 K 412
 DB 429 K 429
 RESULT 33
 ABB49954
 ID ABB49954 standard; Protein; 744 AA.
 XX
 XX ABB49954;
 AC
 XX
 DT 05-FEB-2002 (first entry)
 XX
 XX
 DE *Listeria monocytogenes* protein #2658.
 XX
 KM Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KM vitamin B12; bacterial infection; disease.
 OS *Listeria monocytogenes*.
 XX
 XX WO200177335-A2.
 XX
 XX 18-OCT-2001.
 PD
 XX 11-APR-2001; 2001WO-FR01118.
 PF
 XX 11-APR-2000; 2000FR-0004629.
 PR
 XX (INSP) INST PASTEUR.
 PA
 PI Buchsesser C, Frangeul L, Couve E, Rusniok C, Fajhi H, Dehoux P,
 PI Duesurget O, Chetouiati F, Nedjari H, Glaser P, Kunst F, Cossart P,
 PI Daniela J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA,
 PI Dominguez-Bernal G, Garrido-Garcia P, Tietzer-Martinez A, Amend A,
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L,
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,
 PI Madueno E, De Pablos B, Wehlend J, Kaerst U, Entian K, Hauf J,
 PI Rose M, Voss H;
 PI
 XX WPI; 2002-010914/01.
 DR
 XX
 PT Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment
 PT and prevention of *Listeria* and related bacterial infections, and
 PT related polypeptides -
 PT
 PS Claim 6; SEQ ID No 2659; 192pp; French.
 XX
 CC The present invention relates to the genome sequence of *Listeria*
 CC *monocytogenes* EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in *L.*
 CC *monocytogenes* and related organisms, and for studying genetic

CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of *L. monocytogenes* and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate *L. monocytogenes*-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccine compositions for the treatment or prevention of infections by *L.*
 CC *monocytogenes* and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 744 AA;

Query Match 4.5%; Score 117.5; DB 23; Length 744;

Best Local Similarity 18.9%; Pred. No. 0.0019; Mismatches 167; Indels 223; Gaps 28;

Matches 112; Conservative 90; Mismatches 167; Indels 223; Gaps 28;

43 VLRDRTDILVAGPGSGAIPKFWLVPCAIIMLYAKLSNLSKQALFYAVGTPP 102

220 LFKDADVSLIG-----TVLVLVFLVLYVR-SPLIALIPIL-AVGFAV 261

103 LIFPALFPTVYPLRDVLPTEPADRLQALPPGLGLVALIRNMTFAFYLAELMGSV 162

262 L-----VITPILGLL-----AKEGITYGSGGL-----SI 286

163 MLMLMFGFANEI-----TKIHEAKRFYALF-----GICANISLASGRAIWMASK 208

287 MTVLVLF-GAGTGYCFLIFRSHLTENKRFQAFSEARSGTAGALAL--SGITVMAAL 343

209 LPAISVSEG-----VDPWGISLKLIM--AMTIYS--GLVLMASYMINKNVLDPREFNP 258

344 LILAAEYSGFHNFAVFPFLAIFPMISLTLVPALGIFGRVSPFMPF-----VPTVEM 397

259 EEMOKGKAKKPRMMKDSFLY-----LDRSPYILTLTLVAVGICIMLIEVTMSQK 313

396 ESTRAKKKKKTPRHHKRNFMHIGEMSAKHPRIILITLILIGGI-----FTTQVK 451

314 LQYPMNDYSEFMGNSFWT-----GVSV----- 338

452 YYYDYLSTFEDMPSEGTLLISDHGAGMLAPMEVNVNKSMSLENNVGVASVTPP 511

339 -----LIM-----LFVGNVIRKFG----- 353

512 ERSKGYKTYTLIKDPYENEMADVVKYRAAADKNGNDVYIAGQFATGYDDRAVTEHDEK 571

354 -----WLTGA-----LYTPVWVLTLGIVFPALVIFRNQASGLVAMFGTT 392

572 VIIPVIALIALLCYLSITAMLYLVATVLLSFVAGLGLGVII--HYAMGVBAISGLI 630

393 PL--MLAVVGAIONILSKSTKVALPSTKEMAVYPLDQEKVKGKALIDVAARFGKS 449

631 FLYAVFVVALBEDVIMFIS--SIWKSCKK--PLK-----AITCEVGOT 673

450 GGALIQGLLVICSGISGAMTPYLAIVL--FIIAIVLSATKLNKLFLAQS 498

674 GGAVTSAG-LIAGTFGVLTLPILQLVQFGLITAGLVLDTFIVRPLVPS 724

RESULT 34

ABB48123 standard; Protein; 493 AA.

AC ABB48123;

DT 05-FEB-2002 (first entry)

XX Listeria monocytogenes protein #827.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;

KW vitamin B12; bacterial infection; disease.

XX Listeria monocytogenes.

XX WO200177335-A2.

XX 18-OCT-2001.

XX 11-APR-2001; 2001WO-FR01118.

XX 11-APR-2000; 2000FR-0004629.

XX (INSP) INST PASTEUR.

PI Buchrieser C, Frangoul L, Couve E, Rusnick C, Psini H, Dehoux P,

PI Duseurget O, Chetouani F, Nedjati H, Glaeser P, Kunst F, Cossart P,

PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng B, Vazquez-Boland JA,

PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,

PI Chakraborty T, Dommann E, Hain T, Berche P, Chardit A, Durant L,

PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,

PI Madueno E, De Pablos B, Weiland J, Kaerst U, Entian K, Haut J,

PI Rose M, Voss H,

DR WPI; 2002-010914/01.

PT Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment

PT and prevention of *Listeria* and related bacterial infections, and

PT related polypeptides

XX Claim 6; SEQ ID No 828; 192pp; French.

XX The present invention relates to the genome sequence of *Listeria*

XX *monocytogenes* EGD-e (see ABB48123). The genome sequence and fragments of

XX it are useful for selecting probes and primers for detecting genes in *L.*

XX *monocytogenes* and related organisms, and for studying genetic

XX polymorphisms and other genomes. The present sequence is a protein

XX encoded by the genome sequence of the present invention. Proteins

XX expressed from the genome sequence are useful for raising specific

XX antibodies, identification of *L. monocytogenes* and related organisms, and

XX for biosynthesis and biodegradation, especially biosynthesis of Vitamin

XX B12. The genome sequence and proteins encoded by it are also useful for

XX selecting compounds that regulate gene expression and cell replication

XX and modulate *L. monocytogenes*-related diseases. In addition, the genome

XX sequence and proteins encoded by it are useful in pharmaceutical and

XX vaccine compositions for the treatment or prevention of infections by *L.*

XX *monocytogenes* and related organisms.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 493 AA;

Query Match 4.4%; Score 117; DB 23; Length 493;

Best Local Similarity 20.5%; Pred. No. 0.0012; Mismatches 109; Conservative 75; Mismatches 20; Indels 146; Gaps 24;

Matches 109; Conservative 75; Mismatches 20; Indels 146; Gaps 24;

27 VLPMFMFCITFNTVLRDT-----KDTLVGABSGAIPFIKFWLVPCAI--IF 78

23 VVTMIIAGFVALINQTLTATLPMIMDDLHTTAAGQ-----WLTTFLLTNGIM 72

79 MLTYAKLSNLSKQALF-----YAVGTPFLIFPALPFT-----VITPLRDV 119

73 IPTALLLEKISKTLFTTAMTVFTIGTIIASVAGSPFILLTGRIVQAAGAGIMPLIQT 132

120 LHPTEPADRLQALPPGLGLVA-----IRNMTFAFYVLAELMGSVMLSP 166

133 IFLLILF-PREKRGAMGLMGVIAFAPAIQPTLSGIVDSYDWRVLFILIPAIYIDIL 191

167 MFWGFANEITIKHEAKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISLRL 226

192 AFFGM-KKVVKLTDTKIDF-----LSIVWS--SIGFGALVGFSSAGNDGMDTIVY 240

227 LMAMTIVGVLVMSYMWINKNVLDPREFYNEEMQKKKAKKPRMMKDSFLYIDRSPY 286

```

Db      241 TM---LIVGVVIALFVW-RQVINDPME-----LHVRYV 274
Qy      287 ILLTLVLVAYGICINLIEVTKSOKLQYPNNDYSEPMGNSFMTGVSVLIMLFVG 346
Db      275 FSLSVLT---GSIVTMAMIGAEIVPLVYIGTRGSSALQSGILLPGALIMGIMSPING 330
Qy      347 NVIRKFG--WLTGALVTPVWVLLTGIVFPALVIFRNQASGLVAMEGT--TPIMLANVGA 402
Db      331 IIFDKIGAKWLT---ITGVTLITGITPF-----MFLTMDPIWYIVF-- 371
Qy      403 IONILSKSTKVALFDSTKEMAVYPLDQEKVKGAIDVVAARFGKSGALLQOGLVVIC 462
Db      372 -----YAVRFGISMAMPV-----STAGMNLPHMLNHSGAVNNTIRQ-----IA 413
Qy      463 GSGIANTPYLAVILLFTIAIWLVSATKLN--KLFLAQSALKEQVAGSDS 510
Db      414 GSIG-----TAVLITV-----LTNVTKDMNPGKALMATDPASFAQAQDAS 454

RESULT 35
AAU38247
ID      AAU38247 standard; Protein; 477 AA.
XX
AC      AAU38247;
XX
DT      14-FEB-2002 (first entry)
XX
DE      Salmonella typhi cellular proliferation protein #138.
XX
KW      Antisense; prokaryotic cellular proliferation protein;
XX      antibiotic; antibacterial; drug design.
XX
OS      Salmonella typhi.
XX
PN      WO200170955-A2.
XX
PD      27-SEP-2001.
XX
PF      21-MAR-2001; 2001WO-US09180.
XX
PR      21-MAR-2000; 2000US-191078P.
XX      23-MAY-2000; 2000US-206848P.
XX      26-MAY-2000; 2000US-207727P.
XX      23-OCT-2000; 2000US-242578P.
XX      27-NOV-2000; 2000US-253625P.
XX      22-DEC-2000; 2000US-257931P.
XX      16-FEB-2001; 2001US-269308P.
XX
PA      (ELIT-) ELITRA PHARM INC.
XX
PI      Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
PI      Yamamoto RT, Xu HH;
XX
DR      MPI: 2001-611495/70.
XX      N-PSDB; AAS56106.
XX
PT      New polynucleotides for the identification and development of
PT      antibiotics, comprise sequences of antisense nucleic acids -
XX
PS      Example 3; Seq ID No 13840; 511pp; English.
XX
XX      The invention relates to antisense inhibitors of genes essential to
XX      prokaryotic cellular proliferation, their use in identifying the
XX      genes, their use in the discovery of novel antibiotics, the essential
XX      genes themselves and the encoded proteins. The prokaryotes used are
XX      Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX      pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX      invention is also useful for the identification of potential new targets
XX      for antibiotic development. The antisense nucleic acids can also be used
XX      to identify proteins used in proliferation, to express these proteins,
XX      and to obtain antibodies capable of binding to the expressed proteins.
XX      The proteins can be used to screen compounds in rational drug discovery

```

```

CC      programmes. The antisense nucleic acid sequence is also useful to screen
CC      for homologous nucleic acids which are required for cell proliferation in
CC      a wide variety of organisms. The present sequence represents an
CC      essential prokaryotic cellular proliferation protein.
CC      Note: The sequence data for this patent did not form part
CC      of the printed specification, but was obtained in electronic
CC      format directly from WIPO at
CC      ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence 477 AA;
XX
Query Match 4.4%; Score 116.5; DB 22; Length 477;
Best Local Similarity 20.2%; Pred. No. 0.0013;
Matches 95; Conservative 84; Mismatches 186; Indels 105; Gaps 27;

Qy      69 WLVPICAIIFMLIYAKLSNI-LSKQALFYANGTPPLIFALPPTIYPLRDVLHTEFPAD 127
Db      51 WISVPCLLAFVWMLFSAVAVNLPKVGFNFTDPLFMLTALPSVSGALLXVPY----- 104
Qy      128 RLQALPPLGLGLVAILNNMTPAFAFYVLAELGWSVMSLMFMGFANEITKIHAKRFVAL 187
Db      105 --SKWVP--LFG---GRRWT--AFST-----GIILVPCWLGXA-----VDISTPST 144
Qy      188 FGIGANISILASGRAIVWASKLRASVS-----EGYDPMGISRLLMAMTI 232
Db      145 FII---ISLLCXXGAXFASSM-ANISFFPKQKGALGUNGGLGNNGVSVMLXAPLV 200
Qy      233 V-----SGIVMASYW-WIN-KNVLTDPRFVNPDEMOKKKAKPRM 272
Db      201 VLSIFAARFSGHVGEPDGSQLYLANAMIWPFIAIFTLAMFGMNEI-----ATSKA 254
Qy      273 NKKDSFVLVDSPPYLLTLTLVIA-YGICINLIEVTKSOKLQYPNNDYSEPMGNSF 331
Db      255 SLKEQLPVLRKG-HLMTWSLVLATFGSFIG-FSAGFAMLSKTQFPDVQ-----ILHTAF 307
Qy      332 WTGVVSVLIMLFVGVGNVIRKGMLTGALVTPV-NVLLTGIVFPALVIFRNQASGLVAMEG 390
Db      308 RGPFIGALAR-SAGALISDRIGTRVTLINFLYMAIFSGLLFLTPT-GGVGSGFIATFG 365
Qy      391 TPIMLAVV-----GAIQNILSKSTKVALFDSTKEMAVYPLDQEKVKGAIDVVA 444
Db      366 ---VFLAFLTPAGLSSGSTFQWISVIFRKLTMDRYKAGS---GSDERAMREAAATDTAA 418
Qy      445 -----RFGSGGALLIQOGL---LVICGS-IGAMTPYLAIVLLFTIAIMLV 485
Db      419 LGFISALGAIIGFFIPKAFGSSLALTGSPVGAMKVFLIFYACVITWAV 468

RESULT 36
AAE30467
ID      AAE30467 standard; Protein; 635 AA.
XX
AC      AAE30467;
XX
DT      24-FEB-2003 (first entry)
XX
DE      Haemophilus influenza cytochrome C-type biogenesis protein.
XX
KW      Virulence; veterinary; infection; pneumonia; antimicrobial drug; vaccine;
KW      antibiotic; gene therapy; cytochrome C-type biogenesis protein;
KW      antibacterial.
XX
XX      Haemophilus influenzae.
XX
OS      Haemophilus influenzae.
XX
PN      WO200277020-A2.
XX
PD      03-OCT-2002.
XX
PF      18-MAR-2002; 2002WO-GB01305.
XX
PR      22-MAR-2001; 2001GB-0007234.
XX      23-MAR-2001; 2001GB-0007360.
XX

```


PA (ISIS-) ISIS INNOVATION LTD.
 XX Herbert MA, Deadman ME, Hood DW, Moxon ER;
 PI WPI: 2003-029913/02.
 XX DR N-PSDB; MAD47835.
 XX PT New virulence gene from Hemophilus influenzae, useful for producing
 PT vaccines or antibiotics for preventing or treating pneumonia -
 XX
 PS Claim 4; Page 70-72; 132pp; English.
 CC The present invention relates to Hemophilus influenzae virulence genes
 CC and proteins encoded by them. The microorganisms or the peptides of
 CC the invention are useful for manufacturing a medicament for treating
 CC (veterinary) or preventing a condition associated with H. influenzae
 CC infection, particularly pneumonia or for identifying an antimicrobial
 CC drug. Sequences of the invention are useful in the production of vaccines
 CC or antibiotics to prevent or treat H. influenzae infection. They are
 CC also used in gene therapy. The present sequence is H. influenzae
 CC cytochrome C-type biogenesis protein.
 XX
 SQ Sequence 635 AA;
 Query Match 4.4%; Score 116; DB 24; Length 635;
 Best Local Similarity 19.9%; Pred. No. 0.0022;
 Matches 103; Conservative 46; Mismatches 142; Indels 226; Gaps 21;
 QY 78 FMIIYAKLSNLSKQALFYAVGTPPLIFPFIYIYELRDVLPTEPADRLQALPPGL 137
 DB 8 FLILLATAS-----AFPLAL-VPOFGLFKKPTLI-----NAWPLSYIFLITATLSIGL 56
 QY 138 LG-----LVAILRMTPFAFYVLAEIEMGSMVLSLMEKGA----- 172
 DB 57 LAYFAVADFTLEYVAHAHNSQQLPFFFYVAATWGHEGSMFWLSLMLAFAFNRK 116
 QY 173 NEITKIHAKRFYALFGIGANISL-----ASGR----- 201
 DB 117 NDRFSAQSLSLGLICGPAVFILFYSNPGRIFFPAEGRDLNPMLODVGILFHPPL 176
 QY 202 -----AIWASKLRA-----SVSEGVDPWGISLRLLMAMTVSGVLMA SY----- 242
 DB 177 YGVGVFAVNFAMSLALINYQASARQIARSKRWGLVWMLFTLTGIVLG-AMWVAYELGW 235
 QY 243 --WINKNVLTDPFYNPEEMQKKGAKKPKMNMKDSFLYIDRSPYILLTLVIVAGIC 300
 DB 236 GGMW-----FMDPVE----- 245
 QY 301 INLEVTWKSQKLQYPMNDYSEFMGNFSFMTGVSVLIMLF-VGQNVIRKFGMLTGAL 359
 DB 246 -NASLMPRLGLALHLSLM--ATEKQGVFSYWTTLFSLAFAFSVLGTFPIVRSGLTSVH 302
 QY 360 V-----TPVWVLTGIVEFALVIFRNOASGLVAMEGTPMLMAVVGAIQNLKSTKYA 414
 DB 303 AFALDNRGVYLL-LIFEVLTAL--AFGLFALRA----- 333
 QY 415 LFDSTKEMAYPLDQKQVAKAIDVVAARF-GKSGGALLQGLLYVC----- 462
 DB 334 -----GSSSESAVKFQFISKSGILLNLTITATVSTFLGTFYP 373
 QY 463 -----GSIGAMPYLL-----AVILLFTIAI 482
 DB 374 MLFQAMWMSISVGSPIYNSIFPPIITAILMLVIVL 410
 RESULT 37
 AAG90107
 ID AAG90107 standard; Protein; 410 AA.
 XX AAG90107;
 AC
 XX 26-SEP-2001 (first entry)
 DT
 XX

DE C glutamicum protein fragment SEQ ID NO: 3861.
 XX
 XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KM organic acid synthesis.
 XX
 XX Corynebacterium glutamicum.
 OS
 XX EPI108790-A2.
 XX
 XX 20-JUN-2001.
 XX
 XX 18-DEC-2000; 2000EP-0127688.
 PF
 XX 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159152.
 ER 03-AUG-2000; 2000JP-0280988.
 XX
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 XX WPI: 2001-376931/40.
 DR N-PSDB; AAH65326.
 XX
 PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 PS Claim 17; SEQ ID NO: 3861; 246pp + Sequence listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
 CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 SQ Sequence 410 AA;
 Query Match 4.4%; Score 115.5; DB 22; Length 410;
 Best Local Similarity 22.8%; Pred. No. 0.0013;
 Matches 72; Conservative 47; Mismatches 104; Indels 93; Gaps 16;
 QY 157 ELWGSVMSL--MFGFANEITKI--HEAKRFYA-----LFGIGANISLAAAG-- 200
 DB 20 KVMIALVALSVFTVAMG--GNERTPLLVFVRBGFSSNLFIDLLVFAIVAGVAGLLAAGPL 78
 QY 201 -----RAIYWASKLRASVSEGVDPWGISLRLLMAM-----TVYSGVLVMA SYWINKNV 249
 DB 79 SDRYGRRAVWLPAPLIIALLSGALIASGEETAILAIGRVLSIGSVGMWMTAGSGMIKX-- 136
 QY 250 LIDRPFYNPEEMQKKGAK--PKNMKDSFL-----YIDRSPYILLTLTLV 294
 DB 137 LSSSRF--BPGVTSAGAKRASVLTGFGALGALAGVMAQWLPDQQLAVVLIHTLTL 193
 QY 295 IAVGICINLEVTWKSQKLQYPMNDYSEFMGNFSFMTGV--SVLIMLFVGGNVIRKF 352
 DB 194 ILPPLILITAPETRSALHKTK-----GSFMSDVIVPSALDKRFL--FVVAPI 238
 QY 353 G-WLTGALVTPWVYL-----LTGIVFPALVIFRNOASG----- 384
 DB 239 GPWVGAAFTAYAVLPQQLRDMVSAVAVYSALIALVLTGSGFGIQDFGPOIMGTSKTRGP 298
 QY 385 LVAMFGTTPMLMAVVV 400

Db 299 ILAMFVTVIGMIGAVI 314

RESULT 38
AAB78947
ID AAB78947 standard; Protein; 410 AA.
XX
AC AAB78947;
XX
DT 30-APR-2001 (first entry)
XX
DE C. glutamicum SRT protein sequence SEQ ID NO:154.
XX
XX Corynebacterium glutamicum; stress; resistance; tolerance; SRT;
fine chemical production; organic acid; proteinogenic amino acid;
nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
evolutionary study; environmental hazard; fermentation.
XX
OS Corynebacterium glutamicum.
XX
FN WO200100804-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00922.
XX
PR 25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030429.
PR 01-JUL-1999; 99US-0142692.
PR 08-JUL-1999; 99DE-1031413.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031541.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032914.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99US-0151214.
PR 31-AUG-1999; 99DE-1041382.
XX
PA (BAD1) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Lee H;
Kim H;
XX
DR WPI; 2001-061972/07.
DR N-PSDB; AAF71060.
XX
XX New isolated Corynebacterium glutamicum nucleic acid encoding a stress,
PT tolerance or resistance protein, for production or modulation of
PT production of fine chemicals, such as, e.g. amino acids, lipids,
PT carbohydrates, or enzymes -
XX
XX Claim 20; Page 314-316; 526bp; English.
XX
XX AAF70984 to AAF71133 encode the Corynebacterium glutamicum stress,
resistance, and tolerance (SRT) proteins given in AAB78971 to AAB79020.
XX The C. glutamicum SRT genes (I) can be used in vectors (II) for
XX expression in host cells and production of fine chemicals, such as, an
XX organic acid, a proteinogenic or nonproteinogenic amino acid (preferred),
XX a purine or pyrimidine base, a nucleoside, a nucleotide, a lipid, a
XX saturated or unsaturated fatty acid, a diol, a carbohydrate, an aromatic
XX compound, a vitamin, a cofactor, a polypeptide, or an enzyme. The fine
XX chemical production can be modulated. The presence of (I) or the SRT
XX proteins (III) encoded by them are used for diagnosing the presence
XX or activity of Corynebacterium diptheriae. (I), (II), (III) and host
XX cells containing them can be used to map the genomes of organisms related
XX to C. glutamicum, to identify and localise C. glutamicum sequences of
XX interest, in evolutionary studies, in determination of SRT protein
XX regions required for function, in modulating the SRT protein activity,
XX and in modulating the activity of an SRT pathway. (II) are used to permit

CC C. glutamicum to survive in an environment that is normally
CC environmentally or chemically hazardous to it. (I) and protein molecules
CC encoded by it increase the survival of C. glutamicum to chemical and
CC environmental hazards and provide a means for continued growth and
CC multiplication in large scale fermentative growth conditions. By
CC increasing the growth rate or maintaining a normal growth rate in poor or
CC toxic conditions, the yield, production and/or efficiency or production
CC of fine chemicals from a culture may be increased.
XX
SQ Sequence 410 AA;
Query Match 4.4%; Score 115.5; DB 22; Length 410;
Best Local Similarity 22.8%; Pred. No. 0.0013;
Matches 72; Conservative 47; Mismatches 104; Indels 93; Gaps 16;
OY 157 ELMSQWLSL--MFWGANEITKT--HEAKRPA-----LFGIGANISLISAG-- 200
DB 20 KMWLAVALSVFTVAMG--GNETPPLVFRGSGFNSLFDLLVFAVIGAVGLAAGPL 78
OY 201 -----RAIVWASKLRASVSGVDPMGISLRILMAM-----TIVSGVLWASYWINKNV 249
DB 79 SDRYGRRAVMDPAFLAILLSALIASGEFTAILAIGRVLSGISVGMWTRGSSWIK-- 136
OY 250 LTDPFYNPEEMOKKKGAK--PRANMKDSPL-----YDRSEFYILLTLV 294
DB 137 LSSSRF---EPGVKTSAGAKRASWSLTGFGALAGVMAQMLPLPGQALAVVHLITL 193
OY 295 IAVGICINLIEVTKWSQKLOQYPMNDYSEPMGNFSWTGV--SVLIMLFVGAVIRKF 352
DB 194 ILFPLILITTAPETRQSAHLKTK-----GSFMSDVLVPSALDKRFL--FVYAPI 238
OY 353 G-WLTGALVTPVWVL-----LTGIVFPALVIFRNQASG----- 384
DB 239 GPWVFGAFTAYAVLPBQLDPMVASPVAISHLILVLTGSGFGIQGPQIMGSKTRGP 298
OY 385 LVAMFGTTPMLAVVV 400
DB 299 ILAMFVTVIGMIGAVI 314
XX
XX RESULT 39
XX AAE05852
XX ID AAE05852 standard; Protein; 392 AA.
XX
XX AC AAE05852;
XX
XX DT 24-SEP-2001 (first entry)
XX
XX Pseudomonas stutzeri open reading frame-N (ORF-N) protein.
XX
XX Open reading frame-N; ORF-N; pyridine-2,6-bis (thiocarboxylate);
XX PUTC; environmental remediation; phytoremediation; bioaccumulation;
XX water purification; solution mining mobilisation; immobilisation;
XX detoxification; redox state modifier; metal ion reactivity;
XX carbon tetrachloride; metal.
XX
XX Pseudomonas stutzeri.
XX
XX WO200153309-A1.
XX
XX 26-JUL-2001.
XX
XX 19-JAN-2001; 2001MO-US02386.
XX
XX 20-JAN-2000; 2000US-0177251.
XX
XX (IDAH-) IDAHO RES FOUND INC.
XX (PASZ/) PASZCZYNSKI A.
XX (SEBA/) SEBAT J L.
XX
XX Paszczynski A, Sebat JL, Lewis TA, Crawford RL, Cortese MS;
XX WPI; 2001-465361/50.

DR N-PSDB; AAD11172.

XX New isolated nucleic acids, useful for producing enzymes required to
 PT produce pyridine-2,6-bis (thiocarboxylate), especially useful for
 PT reducing the amount of metal or carbon tetrachloride in a substrate,
 PT e.g. soil or water

XX Claim 32; Page 100-101; 172pp; English.

XX The present sequence is a Pseudomonas stutzeri open reading frame-N
 CC (ORF-N) protein. The Pseudomonas stutzeri genome includes ORFs that
 CC encode enzymes required for biosynthesis of pyridine-2,6-bis
 CC (thiocarboxylate) (PDBC). The ORFs encoding PDBC are especially useful
 CC in environmental remediation methods, e.g. phytoremediation,
 CC bioaccumulation, water purification, waste water purification, solution
 CC mining mobilization, immobilization, detoxification, redox state
 CC modifier or modification of metal ion reactivity. In particular, the
 CC ORFs are useful for degrading carbon tetrachloride and removing metals
 CC from substrates, e.g. soil or water.

XX Sequence 392 AA;

Query Match 4.4%; Score 115; DB 22; Length 392;
 Best Local Similarity 20.9%; Pred. No. 0.0014;
 Matches 88; Conservative 55; Mismatches 127; Indels 152; Gaps 20;

OY 94 LFVANGTPTLIFALPPTVITPLRDVLAHTERAD---RLQ-AIIPGLGLVALIRWT 148
 DB 18 LLIAMGPMWMTFYAI-----GILGPHLVADIGISROQLGWLTASTFGLAALISPA 68
 OY 149 FA-----AFVYLAELMGSVMLSMFMGPNETHKHEAKRFALFGIGANISL 196
 DB 69 GALVQRMGRAGLICMFLVIGLSFSLMANVP--GGGLVTL-----LLCGTAQSLAN 119
 OY 197 LASGAIWASKL-BASVSEGVDPWGISRLMAMTIVSGVLVLM----- 239
 DB 120 PATNOAIASVPAARRAGVGLKQSGVQASALLAGVALPPLVLMGMGALAAWVVALV 179
 OY 240 ----ASYMWINQVLTDRFYFPEEMQKKGAKPRAMKDSFLYLDSPYLLTLTVI 295
 DB 180 MAALVTYWPAAKSV-----SAPSLPVR-----VGGPNVMSILMAI 215
 OY 296 AVGICINILEVTWKSQKLQYNNNDYSEFMGNFSPWTCV--VSVLIMLVGNGVIRKF-- 352
 DB 216 Q--LQAGLA-----LSFMTFLGAVTAQIGSVSTI-----GAMVSCGGA 253
 OY 353 -GMLTGAIVTVM-----VLLTGIVF---PALVIFR--NOASGLVAMFGTTPMLAVV 399
 DB 254 MGIISRVLITPADIKDKETILLGLVFLIAGLALAVMEANTQGMPLMLGVTMGLTV- 312
 OY 400 VGAIQNIISKSTKYALFSDTKEMAYIPLDQEQVKGAIDVVAARFGKSGGALIQGGL 459
 DB 313 --AASNAIAMS-----MLIRDRGFGGAATSAAGML 339
 OY 460 VI 461
 DB 340 SV 341

RESULT 40
 AAG91270
 ID AAG91270 standard; Protein; 708 AA.
 XX AAG91270;
 AC
 XX 26-SEP-2001 (first entry)
 DT
 XX C glutamicum protein fragment SEQ ID NO: 5024.
 DE
 XX Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;
 KM organic acid synthesis.
 XX Corynebacterium glutamicum.
 OS

XX EPI108790-A2.
 PN 20-JUN-2001.
 XX 18-DEC-2000; 2000EP-0127688.
 XX 16-DEC-1999; 98JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX (KYOWA) KYOWA HAKKO KOGYO KK.
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A,
 DR WPI; 2001-376931/40.
 DR N-PSDB; AAH66489.

PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene

PS Claim 17; SEQ ID NO: 5024; 246bp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from corynebacterium bacterium, and identifying a homologue of a gene derived
 CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the amplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX Sequence 708 AA;

Query Match 4.4%; Score 115; DB 22; Length 708;
 Best Local Similarity 18.6%; Pred. No. 0.0034;
 Matches 106; Conservative 70; Mismatches 149; Indels 244; Gaps 25;

OY 109 PPTVITPPLRDVLAHTERADRLQAILPPG-----LGLVALIRMTFAAFVLAEL 158
 DB 120 YPHVPRAPRAKPPMTSEKIMRGVAIGGVITVAGVLLVSAIQGM-----LGPLGRV 175
 OY 159 WGSVWLSIMFMGPNETHKHEAKRFYALF-----GIGA 192
 DB 176 IGAVLLAVLLIGAHHYAK--RGTRVEALVALTTSQIAPLATTSATIFILEMPPGIGS 233
 OY 193 ----NISLASGRAIYMA-SKLASVSEGVDPWGISRLMAMTIVSG-----LVMA 240
 DB 234 LVAILIGNIGFLIVK--LMSLSKTEKSAAG-----HTVFVAIAVSGFSAIPALSA 284
 OY 241 SYMMINKNVLTDRFYFPEEMQKKGAKPRAMKDSFLYLDSPYLLTLTVIAGIC 300
 DB 285 DAWM-----PIRSIVAALLSTYIS 304
 OY 301 INLI-----EVTWKSQ-----LKLQYNNNDY 322
 DB 305 TIIIRASWAPAVILQFVLASQMTWMPATIVGTVALLVALTMDPFXITATDSHDI 364
 OY 323 S--EFMGNF-----SFTGVSUULIMLVGAGNVIRKGMGLTGAIVTVMVLLTGIVFPAL 375
 DB 365 ALEEYMSFETNPVSTWGAVSPLVIVITTSMEIADVMPVLAIPACAVAAIGI--FAL 422
 OY 376 ----VIFRQASGLVAMFG-----TTPMLAV--VVGAA----- 402
 DB 423 RSSDTASIEQMRRLIADVGLALIAETFVQLTFYGDLPNTNLLVWVFLIAGALPMLIRM 482

```
Qy 403 -----IQNLSKSTKY-----ALPDSTKEMAYIPLDQEQ 431
Db 483 LPPORQLGVPPVWVAMLIAAVAMTGVLNRVVISPLMLTDQALIQALLIVFI----- 536
Qy 432 KYGKAIDVVAARFGKSGGALIOGSLVI---CGSIGAMTPYL-----AVILLFI 480
Db 537 ---AATIOVRRSPYGHKLMLOILVGLTLLTSLAISIVTITTFIGRLIAGNAGMMLGFLI 592
Qy 481 A-----IWLVSATKL---NKLFLAQSAL 500
Db 593 GHATVSIILMWVIAAALMLNRKLLDAFGAL 621
```

Search completed: November 25, 2003, 10:14:31
Job time : 59 secs

GenCore Version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 25, 2003, 10:16:48 ; Search time 404 Seconds
(without alignments)
3441.115 Million cell updates/sec

Title: US-09-869-433-2
Perfect score: 2630
Sequence: 1 MTKEEKPFGKLRSEFLWPIH.....AQSAKKEQVQAQEDSAPASS 515

Scoring table:
BLOSUM62
Xgapop 10.0, Ygapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODL=frame+g2n.model -DEV=xlh
-O=/cgn2_1/USPRO.epool/US09869433/runat_24112003.171458.18110/app_query.fasta_1.711
-DB=N Geneseq.19jun03 -QEMT=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=pco -NOR=ext -HEAPSIZE=500 MINLEN=0 -MAXLEN=2000000000
-USER=US09869433 @CGN_1_1_0 @runat_24112003.171458.18110 -NCPU=6 -ICPU=3
-NO MMAP -LARGESUBSTRY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.Geneseq.19jun03.*

1: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA2001.DAT.*
23: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA2002.DAT.*
24: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA2003.DAT.*
25: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|---------|-------|-------------|
| 1 | 2630 | 100.0 | 1637 | 21 | AAA30922 |
| 2 | 2617 | 99.5 | 273254 | 21 | AA081914 |
| 3 | 2617 | 99.5 | 123025 | 20 | AA091990 |
| 4 | 2141 | 81.4 | 1038602 | 20 | AA201425 |
| 5 | 1333 | 50.7 | 1875 | 24 | AB212889 |
| 6 | 1330.5 | 50.6 | 2146 | 21 | AA36046 |
| 7 | 1306 | 49.7 | 1896 | 22 | AA61374 |
| 8 | 1260 | 47.9 | 1770 | 22 | AA61373 |
| 9 | 1025.5 | 39.0 | 1823 | 22 | AA68446 |
| 10 | 1025.5 | 39.0 | 1230025 | 20 | AA091990 |
| 11 | 256 | 9.7 | 284 | 24 | ABL74862 |
| 12 | 242 | 9.2 | 246 | 24 | ABL76356 |
| 13 | 242 | 9.2 | 280 | 25 | ABX86909 |
| 14 | 227.5 | 8.7 | 309 | 24 | ABL71938 |
| 15 | 168 | 6.4 | 265 | 25 | ABX84884 |
| 16 | 162 | 6.2 | 1614 | 22 | AA65082 |
| 17 | 162 | 6.2 | 349980 | 22 | AA64966 |
| 18 | 152.5 | 5.8 | 1038602 | 20 | AA201425 |
| 19 | 144 | 5.5 | 1983 | 21 | AA253363 |
| 20 | 144 | 5.5 | 1983 | 21 | AA253364 |
| 21 | 144 | 5.5 | 2028 | 24 | ABK54090 |
| 22 | 144 | 5.5 | 60873 | 21 | AA61469 |
| 23 | 144 | 5.5 | 349980 | 21 | AA621610 |
| 24 | 144 | 5.5 | 1437668 | 21 | AA681490 |
| 25 | 138.5 | 5.3 | 1389 | 23 | AA552427 |
| 26 | 137 | 5.2 | 1613 | 24 | AB070453 |
| 27 | 134.5 | 5.1 | 349980 | 24 | AB081848 |
| 28 | 134 | 5.1 | 1983 | 21 | AA253362 |
| 29 | 134 | 5.1 | 2025 | 25 | AB239606 |
| 30 | 134 | 5.1 | 349980 | 22 | AA64966 |
| 31 | 133 | 5.1 | 1830121 | 17 | AA742063 |
| 32 | 132.5 | 5.0 | 1392 | 23 | AA552410 |
| 33 | 129.5 | 4.9 | 294458 | 24 | AB403041 |
| 34 | 126 | 4.8 | 664707 | 24 | AB067196 |
| 35 | 126 | 4.8 | 3011208 | 24 | AB069245 |
| 36 | 125.5 | 4.8 | 1305 | 25 | ABT15037 |
| 37 | 123.5 | 4.7 | 7471 | 21 | AA288789 |
| 38 | 123 | 4.7 | 2365869 | 24 | AB490521 |
| 39 | 122.5 | 4.7 | 68750 | 21 | AA255887 |
| 40 | 122 | 4.6 | 2547 | 23 | ABL17965 |
| 41 | 122 | 4.6 | 2564 | 23 | ABL18561 |
| 42 | 121 | 4.6 | 2541 | 24 | AB070945 |
| 43 | 121 | 4.6 | 5482 | 24 | ABK12653 |
| 44 | 120.5 | 4.6 | 1188 | 23 | AA56157 |
| 45 | 120.5 | 4.6 | 1278 | 23 | AA554436 |

ALIGNMENTS

| | |
|----------|---|
| RESULT 1 | |
| AAA30922 | AAA30922 standard; DNA; 1637 BP. |
| XX | XX |
| AC | AAA30922; |
| XX | XX |
| DT | 22-SEP-2000 (first entry) |
| XX | XX |
| DE | C. pneumoniae ATP/ADP translocase coding sequence. |
| XX | XX |
| KW | ATP/ADP translocase; Chlamydia infection; diagnosis; therapy; ds. |
| OS | Chlamydia pneumoniae. |
| XX | XX |
| FT | Key |
| FT | CDS |
| FT | Location/Qualifiers |
| FT | 51..1598 |
| FT | /*tag= a |

FT /product= ATP/ADP_translocase
 XX WO200039157-A1.
 XX 06-JUL-2000.
 XX 22-DEC-1999; 99WO-CA01224.
 XX 28-DEC-1998; 98US-0114060.
 XX 12-MAR-1999; 99US-0123967.
 XX 30-JUN-1999; 99US-0141271.
 XX (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Murdin AD, Oomen RP, Wang J, Dunn P;
 XX WPI, 2000-452368/39.
 DR P-PSDB; AAV90265.
 XX
 PT Novel Chlamydia polynucleotides and polypeptides, useful for diagnosis,
 PT prevention and treatment of Chlamydia infection in mammals -
 XX
 PS Claim 2, Fig 1, 81pp; English.
 CC This sequence encodes the Chlamydia pneumoniae ATP/ADP translocase
 CC of the invention. The protein, DNA encoding it, or a vaccine containing
 CC the DNA or protein, are useful for diagnosing, preventing or treating
 CC Chlamydia infection. The sequences can also be used in a method for
 CC the detection of Chlamydia infection. Primers or probes derived from the
 CC DNA sequence are useful in diagnostic tests for detecting Chlamydia
 CC infection.
 XX
 SQ Sequence 1637 BP, 431 A, 327 C, 325 G, 554 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 1.81e-276 Length: 1637
 Score: 2630.00 Matches: 515
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0
 US-09-869-433-2 (1-515) x AAA03922 (1-1637)
 QY 1 MetThrLysThrGluGluLysProPheGlyLysLeuArgSerPheLeuTrpProIleHis 20
 Db 51 ATGCAAAAGCGAAGAAACCTTTTGGAAATGCGCTCTTCTTGCGCGATACAT 110
 QY 21 ThrHisGluLeuLysValLeuProMetPheLeuMetPhePheCysIleThrPheAsn 40
 Db 111 ACTCAGAGCTBAAGAAAGTCTGCGCAATGTTCTTAATGTTCTTCTGATTAATTTTAAAC 170
 QY 41 TyrThrValLeuArgAspThrLysAspThrLysLeuIleValGlyAlaProGlySerGlyVala 60
 Db 171 TATACGGGTGTTACGGATCAAAAGACATCTTATTTGGGAGCGCTCGGTTCGGTCA 230
 QY 61 GluAlaIleProPheIleLysPheTrpLeuValValProCysAlaIleIlePheMetLeu 80
 Db 231 GAGGGAATACCTTCATCAAGTTTGGCTTGTGCTGCTGCTGCTATATATCTTATGCTT 290
 QY 81 IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr 100
 Db 291 ATTATGCAAGAGCTAAGTAATTTTAAAGAGAGGCGCTTATTTATGACAGTGGAAAG 350
 QY 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
 Db 351 CCGCTTTTATTTCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 410
 QY 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140
 Db 411 CATCTCTCAAGAAATTTGCTGACCGCTTACAGGCGCATCTTACCGAGGATTCCTAGGACTC 470
 QY 141 ValAlaIleLeuArgAsnTrpThrPheAlaIlePheTyrValLeuAlaGluLeuTrpGly 160

Db 471 GTTGCAATCTTAAGAACTGACATTTGCTGCACTTTATGACTTGCTGACATGAGGGA 530
 QY 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleHisGlu 180
 Db 531 AGCGTCATGCTATCTTAATGTTCTGGGGAATTTCTTAATGAATTAACAAATCCACCAA 590
 QY 181 AlAlaValArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
 Db 591 GCAAGGCTTTCTACGCTCTTTCGTTACGAGCTAATATTTCTTACAGCTTCTGCT 650
 QY 201 ArgAlaIleValITrPalSerLysLeuArgAlaSerValSerGluGlyValAspProTrp 220
 Db 651 CGTGCAATTTGTTGGCTTCGCAAGTGAAGTCCGTTCTGAGGATGATCCCTGG 710
 QY 221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla 240
 Db 711 GGAATTTCTTACCTGTTTGTGATGCTATGACATATGATCTGAGACTGTTCTTATGCGC 770
 QY 241 SerTyrTrpTrpIleAsnLysAsnValLeuThrAspProArgPheTyrAsnProGluGlu 260
 Db 771 AGTACTGCTGATCAATGAAGACGATGACCGATCCTGCTCTTAATCCAGAAAGAA 830
 QY 261 MetGlnLysGlyLysLysGlyAlaLysProLysMetLysAspSerPheLeuTyr 280
 Db 831 ATGCAAAAGGGAAGAAAGGCTTAACCTTAATGAATGAAGATAGCTTCTCTAT 890
 QY 281 LeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCys 300
 Db 891 CTGATATGATCTCCTTATATCTTTTATTAATCTCTGTTATGCTATGATTTGCT 950
 QY 301 IleAsnLeuIleGluValIThrTrpLysSerGluLeuLysGlnTyrProAsnMetAsn 320
 Db 951 ATTAACTTAATCGAAGTACTGGAAGTGAAGTGAACCTGAATATCTTAATATGAT 1010
 QY 321 AspTyrSerGluPheMetGlyAsnPheSerPheTrpThrGlyValAlaSerValLeuIle 340
 Db 1011 GACTATAGTATGATTCATGAGGGAATCTTCTCTGAGCTGCGATGATCGTATATC 1070
 QY 341 MetLeuPheValGlyIleAsnValIleArgLysPheGlyTyrTrpLeuThrGlyAlaLeuVal 360
 Db 1071 ATGCTAATTTGTTGTTGTTAACTGTCATTCGTAATTTGATGTTAACTGAGCGCTTATGTC 1130
 QY 361 ThrProValMetValIleLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn 380
 Db 1131 ACTCTGTCAATGTTCTCTCAACAGTATGTTTCTGCTCTGTTATCTTTAGAAAC 1190
 QY 381 GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValAlaVal 400
 Db 1191 CAAGCTTCTGGGCTGCTGCTGCTATGTTGCTGCAACTCTCTCATGCTAGCTGCTGCTGTC 1250
 QY 401 GlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLys 420
 Db 1251 GAGACTTATCAAGAAATATCTTTCGAAATCCAAATCAACCTCTCTTGTGCTCAACCTAAA 1310
 QY 421 GluMetAlaTyrIleProLeuAspGlnGluLysValIleGlyLysAlaIleAsp 440
 Db 1311 GAAATGGCTATATCTCTTTCGACCAAGACCAAAATCAAGGTAAGGCTGCTATGAT 1370
 QY 441 ValValAlaIleArgPheGlyLysSerGlyGlyAlaLeuIleGlnGlnGlyLeuVal 460
 Db 1371 GTAGTTCGCCGCCCTTCGGAATAATCGAGAGAGCTTATCCACAAGGTTTCTGCT 1430
 QY 461 IleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValIleLeuLeuPheIle 480
 Db 1431 ATCTGTGAAGTATTTGAGCTATGACCCCTTATCTTGCACTGATCTTCTTTCATCAT 1490
 QY 481 AlaIleThrPheValSerAlaThrLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeu 500
 Db 1491 GCTATTTGTTGTTCTTCTGCACTTAAGTAAACAACTATTTCTTAAAGGCGAGCTGCTCT 1550
 QY 501 LysGluGlnGluValAlaGlnGlnLysSerAlaProAlaSerSer 515

DB 1551 AAGAACACGAAGTGGCTCAAGAAGTTGAGCTCCGTCTTCTTA 1595

RESULT 2
ID AAC81914/c standard; DNA; 273254 BP.
XX AC AAC81914;
XX DT 27-FEB-2001 (first entry)
XX DB Chlamydia pneumoniae genome DNA.
XX KW Genome; diagnosis; vaccine; ds.
XX OS Chlamydia pneumoniae.
XX PN WO200027994-A2.
XX PD 18-MAY-2000.
XX PF 12-NOV-1999; 99WO-US26923.
XX PR 12-NOV-1999; 98US-0108279.
XX PR 08-APR-1999; 99US-0128606.
XX XX (REGC) UNIV CALIFORNIA.
XX PA Stephens R, Mitchell W, Kaiman S, Davis R,
XX PI MPI, 2000-376516/32.
XX DR
XX PT Isolated nucleic acid for use in diagnostic and analytical methods
XX PT encodes genomic sequence of Chlamydia pneumoniae -
XX PS Claim 2; Page 128-320; 320pp; English.

CC This invention describes a novel nucleic acid (N1) encoding a Chlamydia
CC pneumoniae protein (PI), given in the specification. The isolated nucleic
CC acid is useful for diagnostic and analytical methods, such as,
CC hybridization-based assays or amplification-based assays. The protein may
CC be used for diagnostic purposes, for their enzymatic or structural
CC activity, or as a vaccine. The invention also describes (1) a probe
CC comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
CC (N2) that hybridizes under stringent conditions to N1; (3) an expression
CC cassette comprising N1 under the transcriptional regulation of a
CC transcriptional initiation region functional in an expression host, and a
CC transcripiional termination region; (4) a cell comprising an expression
CC cassette of (3) as part of an extrachromosomal element or integrated into
CC the genome of a host cell as a result of induction of the expression
CC cassette into the host cell; and the cellular progeny of the host cell;
CC (5) a method for producing a PI comprising growing a cell of (4) where
CC the protein is expressed and isolating the protein free of other
CC proteins; (6) a purified polypeptide composition comprising at least 50
CC weight % of PI; and (7) a monoclonal antibody binding specifically to the
CC peptide of (6).

SQ Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;

Alignment Scores:
Pred. No.: 8,43e-272 Length: 273254
Score: 2617.00 Matches: 513
Percent Similarity: 99.61% Conservative: 0
Best Local Similarity: 99.61% Mismatches: 2
Query Match: 99.51% Indels: 0
DB: 21 Gaps: 0

UY US-09-869-433-2 (1-515) x AAC81914 (1-273254)

UY 1 MetThrIlySThGlUluYsPProPhcGIYlYsLeuAArgSerPheLeuTrpProlleHis 20
DB 212420 ATGCACAAATAACCGAAGAAAACCTTTTGAAAAATTTCGCTCTTTCTTGCGCGAATCAT 212361
21 ThrHISglULeuILysValLeuPProMet PheLeuMet PheDncCySilthrPheAsn 40

| | | | |
|----|--------|--|--------|
| Db | 212360 | ACTCAGAGCTTAAGAAAAGTTCTGCCAATGTTCCCTTAATGTTCTTCTGTAATTACATTTAAC | 212301 |
| QY | 41 | TYTTHValLeuArgAspThrLysAspThrLeuIleValAlaProGlyLeuGlyAla60 | |
| Db | 212300 | TATACGGTGTTACCGGATACAAAAGACACTTTATGTGGAGCTCTCGTTCTGGTGC | 212241 |
| QY | 61 | GIuAlaIleProPheIleLysPheTrpLeuValProCysAlaIleIlePheMetLeu80 | |
| Db | 212240 | GAGGCAATACCTTTCATCAGAAGTTTGGCTGTGTGCCCCCTGGCTAATATCTTATGCTT | 212181 |
| QY | 81 | IleTryAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTryAlaValGlyThr100 | |
| Db | 212180 | ATTATGCAAGCTTAGTAATATTTTAAGTAGCAGGCTTATTTTATTCAGAGTGGAAAG | 212121 |
| QY | 101 | ProPheLeuIlePhePheAlaLeuPheProThrValIleTryProLeuArgAspValLeu120 | |
| Db | 212120 | CCCTTTTAATTTTCTTTGGCCCTGTTCGCACTGTATATTTATCCGCTACGCAATGTTTTA | 212061 |
| QY | 121 | HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu140 | |
| Db | 212060 | CATCCCTACAGAAATTTGCTGACCGGTGTACAGGCCATCTACCTCCAGGAATGTGACGACTC | 212001 |
| QY | 141 | ValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTryValLeuAlaGluLeuTrpGly160 | |
| Db | 212000 | GTTGCGATCTTAGAAACTGGACATTTGCTGATTTTAATGACTGTGCTAAGCTATGAGGA | 211941 |
| QY | 161 | SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleHisGlu180 | |
| Db | 211940 | AGCGTCATGCTATCTCTAATGTTCTGGGGAATTTGCTAATGAATTAACAANAATCCAGAA | 211881 |
| QY | 181 | AlaLysArgPheTryAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly200 | |
| Db | 211880 | GCAAAAGCTTTCACGCTCTTTCCGTAATCCGATCAATATTTCTTACTAGCTTCGGT | 211821 |
| QY | 201 | ArgAlaIleValTrpAlaSerLysLeuArgAlaSerValSerGluGlyValAspProTrp220 | |
| Db | 211820 | CGTGAAATGTTTGGGCTTCAAGTGTGAGAGCTCCGTTTCGAAGGTGTAGATCCCTGG | 211761 |
| QY | 221 | GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla240 | |
| Db | 211760 | GGAATTTCTTACGCTCTTTGATGGCTATGACTATTTGATTCGGAATTTCTTATAGGCC | 211701 |
| QY | 241 | SerTryTrpTrpIleAsnLysAsnValLeuThrAspProArgPheTryAsnProGluGlu260 | |
| Db | 211700 | AGTACTGCTGGATCAATAGAAGCTATGACCGATCCCTCGCTTCTATATCCAGAAAGA | 211641 |
| QY | 261 | MetGlnLysGlyLysLysGlyValAlaLysProLysMetLysMetLysAspSerPheLeuTrp280 | |
| Db | 211640 | ATGCAAAAGGGGAAAAAAGGCTTAACCTAAAGAAATATAGAAATATGCTTCCTCTAT | 211581 |
| QY | 281 | LeuAspArgSerProTryIleLeuLeuLeuThrLeuLeuValIleAlaTryGlyIleCys300 | |
| Db | 211580 | CTTGCTGATCTCTCTTATATCTTTTATTAAGCTCTCTTGAGTTATTCCTTAATGATTTGC | 211521 |
| QY | 301 | IleAsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGlnTryProAsnMetAsn320 | |
| Db | 211520 | ATTAACTTAATCGAAGAGCACTTGGAAAAAGTCAGTGAACCTGCAATATCTTAATAGAAAT | 211461 |
| QY | 321 | AspTrySerGluPheMetGlyAsnPheSerPheTrpTrpGlyValAlaSerValLeuIle340 | |
| Db | 211460 | GACTATATGATGATTCATGAGGGAACTTCCCTTCGACGCGGTGTATCCGTAAGCTATATC | 211401 |
| QY | 341 | MetLeuPheValAlaGlyAsnValIleArgLysPheGlyTrpLeuThrGlyAlaLeuVal360 | |
| Db | 211400 | ATGCTATTATTGTTGGTGTAACGCTAATTCGTAAATTTGGATGGATTAATCGAGCCCTTAGTC | 211341 |
| QY | 361 | ThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn380 | |
| Db | 211340 | ACTCGTGCATAGGTTCTCTCTAACAAGATATCGTTTTCTTGCGTCTTGTTATCTTTAGAAAC | 211281 |
| QY | 381 | GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValValVal400 | |

Db 211280 CAAGCTTCTGGGCTGTCGATGTTGGTACAACTCCTCATGCTAGCTGTGTC 211221
 QY 401 GYAlaAlleglnaenlleuSerLysSerThrlyTyAlaLeuPheAspSerThrLys 420
 Db 211220 GGAGCTATACGAATATCTTCGAAATCCAAATACGCTCTTGGACTCACTAA 211161
 QY 421 GlnuwerlaTyrllePLeuAspGlnGlnGlnValLysGlyLysAlaAlaLeuAsp 440
 Db 211160 GAATGGCTATATCCCTTCGACCAAGCAAAAGTCAAAGGTAAAGGCTGATATGAT 211101
 QY 441 ValValAlaAlaArgPheGlyLysSerGlyAlaLeuIleGlnGlnGlyLeuLeuVal 460
 Db 211100 GTATTGCTGCGCCCTTCGAAAAATCAGAGAGCTTTAATCCAAACAAGTTGCTGCTT 211041
 QY 461 IleGlySerIleGlyAlaMetThrProTyrLeuAlaValIleLeuPheIleIle 480
 Db 211040 ATCTGTGAAGTATGTGAGCTATGACCCCTTATCTTGCAGTGAATCTTTCATCAT 210981
 QY 481 AlaIleThrLeuValSerAlaThrLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeu 500
 Db 210980 GCTATTTGGTGTGTTTCTGCACTTAAGTTAAACAAATATCTTAGCGCAGCTGTCTT 210921
 QY 501 LysGlnGlnGlnValAlaGlnGlnAspSerAlaProAlaSerSer 515
 Db 210920 AAGAACAAGAGTGGCTCAAGAAAGATTCACTCCTGCTTCTCA 210876
 RESULT 3
 AAX9190
 ID AAX9190 standard; DNA; 1230025 BP.
 XX AAX9190;
 AC AAX9190;
 XX 13-SEP-1999 (first entry)
 DT 13-SEP-1999 (first entry)
 XX Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
 DE Respiratory disease: pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KM vaccine; neutralising epitope; ss.
 XX Chlamydia pneumoniae.
 OS Chlamydia pneumoniae.
 XX MO9927105-A2.
 PN 03-JUN-1999.
 PD 20-NOV-1998; 98WO-IB01890.
 PF 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 PA (GENSET) GENSET.
 PI Griffiths R;
 XX WPI; 1999-357842/30.
 DR Genome sequence of Chlamydia pneumoniae
 PT Claim 1; Page 291-611; 1912pp; English.
 PS The present sequence represents the complete genome of Chlamydia
 CC pneumoniae, and encodes proteins AY34584-Y35879. C. pneumoniae causes
 CC respiratory disease such as pneumonia and bronchitis and is thought
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
 CC encoded by the open reading frames of the C. pneumoniae genome (see
 CC AY34584-Y35879) can be used in immunogenic compositions as vaccines.
 CC Vectors containing C. pneumoniae nucleotide sequences can also be
 CC used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae.
 XX Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Alignment Scores:
 Pred. No.: 7,61e-271 Length: 1230025
 Score: 2617.00 Matches: 513
 Percent Similarity: 99.61% Conservative: 0
 Best Local Similarity: 99.61% Mismatches: 2
 Query Match: 99.51% Indels: 0
 DB: 20 Gaps: 0
 US-09-869-433-2 (1-515) x AAX9190 (1-1230025)
 QY 1 MetThrLysThrGlnGlnGlnLysProPheGlyLysLeuArgSerPheLeuTyrProIleHis 20
 Db 404370 ATGACAAAACCAAGAAAACCTTTGAAAATGGCGCTCTTGTGTGGCCGATCAT 404429
 QY 21 ThrHisGlnLeuLysLysValLeuProMetPheLeuMetPheCysIleThrPheAsn 40
 Db 404430 ACTCACGAGCTAAAGAAAGTCTGCCAATGTCTTAATGTCTTGTATTAAC 404489
 QY 41 TyrThrValLeuArgAspThrLysAspThrLeuIleAlaGlyAlaProGlySerGlyAla 60
 Db 404490 TATACGGTGTACGCGATACAAAAGACATCTTATGTGAGAGCTCTGTGTTCTGAGCA 404549
 QY 61 GluAlaIleProPheIleLysPheTyrLeuValValProCysAlaIleIlePheMetLeu 80
 Db 404550 GAGGCAATACCTTTCATCAAGTTTGGCTGTGTGCTCCCGTGTATATCTTATGCTT 404609
 QY 81 IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr 100
 Db 404610 ATTTATGCAAGCTTAAGTAAATATTTTAAGTACAGGCTTATTTATGACGTGGAAACG 404669
 QY 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
 Db 404670 CCTTTTAAATTTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 404729
 QY 121 HisProThrGlnPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuGlyLeu 140
 Db 404730 CATCTTCACAAATTTGCTGACCGTTTACAGCCATCTTACTCCAGATTGCTAGACATC 404789
 QY 141 ValAlaIleLeuArgAsnTyrPheAlaAlaPheTyrValLeuAlaGlnLeuTyrGly 160
 Db 404790 GTTGCCATCTTAAGAAACGAGACATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 404849
 QY 161 SerValMetLeuSerLeuMetPheTyrGlyPheAlaAsnGlnIleThrLysIleHisGly 180
 Db 404850 AGGTCATGCTATCTTAAGTCTGAGGATTTGCTATATAAATACAAATAATCCAGAA 404909
 QY 181 AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
 Db 404910 GCAAAGGCTTCTACGCTCTTTCGGATCGAGCTAATATTTCTTACAGCTTCGT 404969
 QY 201 ArgAlaIleValTyrPalaserysLeuArgAlaSerValSerGlnGlyValAspProTyr 220
 Db 404970 CGTGCAATTTGTTGGGCTTCAAGTTCAGAGCTTCCCTTCTCAAGGTGAGATCCCTTGG 405029
 QY 221 GlyLysSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla 240
 Db 405030 GGAATTTCTTACGCTTTTGAAGGCTATGACTATGTATCTGAGACTGTCTTATAGCC 405089
 QY 241 SerTyrTrpTyrIleAsnLysAsnValLeuThrAspProArgPheTyrAsnProGlnGlu 260
 Db 405090 AGTTTACGTGGATCAATTAAGAACGTATTAACGATCTCTGCTTCTATATCCAGAA 405149
 QY 261 MetGlnLysGlyLysLysGlyAlaLysProLysMetAsnMetLysAspSerPheLeuTyr 280
 Db 405150 ATCAAAAGGAGGAAAAAGGTGCTAAACCTAAATGAATGAAGATACCTTCTCTAT 405209
 QY 281 LeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCys 300
 Db 405210 CTGCTAGACCTCTTATATCTTATATGCTCTCTGTTATGCTTATGAGATTTGC 405269
 QY 301 IleAsnLeuIleGluValThrTyrLysSerGlnLeuLysLeuGlnTyrProAsnMetAsn 320


```

Db 405270 ATTAACCTTAATCGAAGTACTGGAAAAAGTCAGCTGAACCTGAAATATCTATATGAAAT 405329
Qy 321 AsPTyTserGluPheMetGlyAsnPheserPheThrPhgGlyValValSerValLeuIle 340
Db 405330 GACTATATGATGATTCAGGGGAACTCTCTCTGAGCTGGCGGATGATTCGATCTATATC 405389
Qy 341 MetLeuPheValGlyGlyAsnValIleArgPhePheGlyThrPheThrGlyValValLeuVal 360
Db 405390 ATGCTATTTGTGGTGTGATACGATTCGTAATTTGGATGCTTAAGTACGAGCCCTTGTC 405449
Qy 361 ThrProValMetValLeuLeuThrGlyIleValPhePheValLeuValIlePheArgAsn 380
Db 405450 ACTGCTCATGATGCTTCTCTTCAACAGATGCTTTCTTGCTGCTTGTATCTTAAAGAAC 405509
Qy 381 GluAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValValVal 400
Db 405510 CAAGCTTCTGGGCTGGTGGCTATGTCGTAACAACTCTCATGCTAGCTGTGTGTGTC 405569
Qy 401 GlyAlaIleGlnAsnIleLeuSerLeuSerLeuThrThrThrValAlaLeuPheAspSerThrIys 420
Db 405570 GGAGCTATACGAAATATCTTTTCAAAATCCAAATAACGCTCTCTTGACTCACTAA 405629
Qy 421 GluMetAlaTyrIleProLeuAspGlnGlnIleValIysGlyValAlaAlaIleAsp 440
Db 405630 GAATGGCTATATCCCTCTTGACCAAGAGCAAAAGTCAAGGTAAAGGCTGCTATGAT 405689
Qy 441 ValValAlaAlaArgPheGlyIysSerGlyGlyValAlaLeuIleGlnGlnIleValLeuVal 460
Db 405690 GTAGTTCGCCGCCCTCGGAAATCAGAGAGACCTTAAATCCAAAGGTTGCTGCTT 405749
Qy 461 IleCysGlySerIleGlyAlaMetThrProTyrLeuValAlaIleLeuPheIleIle 480
Db 405750 ATCTGTGAAGATTTGAAGCTATGACCCCTTATCTTGAGTATCTTCTTTCATCAT 405809
Qy 481 AlaIleTyrLeuValSerAlaThrIysLeuAsnIysLeuPheLeuAlaGlnSerAlaLeu 500
Db 405810 GCATATTGGTTGGTTCTTGCAACTAAAGTAAACAATCTATTAGGCGAGTGTGCTT 405869
Qy 501 LysGlnGlnIleValAlaGlnIleAspSerAlaProAlaSerSer 515
Db 405870 AAGAAACAAAGAGTGCCTCAAGAAATGACCTCTCTCTCTTCA 405914

```

RESULT 4

AAZ01425
ID AAZ01425 standard; DNA, 1038602 BP.

XX AAZ01425;

DT 07-OCT-1999 (first entry)

DE Complete genome sequence of Chlamydia trachomatis.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KM paratrachoma; inclusion conjunctivitis; genital disease; perithelphalitis;
 KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KM bartolinitis; pneumopathy; venereal lymphogranulomatosis; ss.
 XX Chlamydia trachomatis.

XX OS W09928475-A2.

XX PD 10-JUN-1999.

XX PF 27-NOV-1998; 98MO-IB01939.

XX PR 04-NOV-1998; 98US-0107077.

XX PR 28-NOV-1997; 97FR-0015041.

XX PR 17-DEC-1997; 97FR-0016034.

XX PA (BEST) GENSET.

XX PI Griffiths R;

```

DR WPI; 1999-371125/31.
XX
PT Genome sequence of Chlamydia trachomatis
XX
PS Claim 1; Page 373-656; 1755pp; English.
XX
CC The present sequence represents the complete genome of Chlamydia
CC trachomatis. Open reading frames (ORFs) of the genome encode
CC polypeptides AY36754-X37949. The polypeptides can be used as vaccines
CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
CC be used to control growth of the microorganism. Chlamydia trachomatis is
CC responsible for a large number of diseases, e.g. eye diseases such as
CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
CC conjunctivitis; genital diseases such as nongonococcal urethritis;
CC epididymitis, cervicitis, salpingitis, perithelphalitis, bartolinitis;
CC pneumopathy in breast feeding infants; and venereal
CC lymphogranulomatosis. The polypeptides of the invention may be of use in
CC treating these diseases.
XX
SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

```

Alignment Scores:

| Pred. No.: | 4,77e-219 | Length: | 1038602 |
|------------------------|-----------|---------------|---------|
| Score: | 2141.00 | Matches: | 409 |
| Percent Similarity: | 87.81% | Conservative: | 52 |
| Best Local Similarity: | 77.90% | Mismatches: | 50 |
| Query Match: | 81.41% | Indels: | 14 |
| DB: | 20 | Gaps: | 3 |

US-09-869-433-2 (1-515) x AAZ01425 (1-1038602)

Qy 1 MetThrIysThrGlnGluIysProPheGlyIysLeuArgSerPheLeuThrProIleHis 20

Db 392407 ATGACTCAACCCCGGAAAAACCTTTGGAAATGGCGCTTCTTCCGCGATACAC 392466

Qy 21 ThrHisGluLeuIysIysValIleProMetPheLeuMetPhePheCysIleThrPheAsn 40

Db 392467 ATGCATGAGCTGAAGAAAGTTCGCCAATGTTCTTAATGTTCTTCTGATATTCAT 392526

Qy 41 TyrThrValLeuArgAspThrIysAspThrLeuIleValGlyAlaProGlySerGlyAla 60

Db 392527 TACACGATTTTGAAGATACAAAAGATACCTTATGCTTACAGACCGGATCTGGAGCA 392586

Qy 61 GluAlaIleProPheIleIysPheThrPheValIleProCysAlaIleIlePheMetLeu 80

Db 392587 GAGGCCATTCCTTCAATTAAGTGTGCTAGTGTGCTCTCTCTCTGTTGCTTCATGCTG 392646

Qy 81 IleTyrAlaIysIysSerAsnIleLeuSerIysGlnAlaLeuPheTyrAlaIleGlyThr 100

Db 392647 ATCTACGCCAAGCTTACCAATATTTGACAAACAGGCTTTTCTTGCGAGGCTCTCA 392706

Qy 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120

Db 392707 CCAATGCTGTATTTCTTCCGACGCTTCCGCTGTATACCTTGCCTGATATTCCT 392766

Qy 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140

Db 392767 CACCCAACAGCTTTCGCTGATCTACATCAATCATCTTCCGCGATTTATGCGGATTC 392826

Qy 141 ValAlaIleLeuArgAsnThrPhePheAlaPheTyrValLeuAlaGlnLeuThrGly 160

Db 392827 ATTCGATGCTACGCACTGACATTTGCTGTGTCTACAGCTTCTTGAACCTTGGGGA 392886

Qy 161 SerValMetLeuSerLeuMetPheThrGlyPheAlaAsnGluIleThrIysIleHisGlu 180

Db 392887 AGCGTTATGCTCTCTTGAAGTCTGGGCTTTCGCCAATTAATTAATAATTAGGGA 392946

Qy 181 AlaIysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200

Db 392947 GCTAAACGTTTCACGCTCTTTCGAGAGGAGCTAACCTACCTCTTAAATTTCTGCT 393006

Qy 201 ArgAlaIleValTyrAlaSerIysLeuArgAlaSerValSerGluIysValAspProTyr 220

```

Db 393007 CCAGCTATTATCTGGCTTTCTTAATTACGCTAGCCTTGGGGAAGCGCTGACCCATG 393066
Qy 221 GlyIleSerIleuValGleuMetAlaMetThrIleValSerGlyValLeuMetAla 240
Db 393067 GGAGTCAGCCCTCTATTTCTCTAATGCGTATGTTCTTGCTTGGCTATTATATGCTGCT 393126
Qy 241 SerTyrTPRIleAsnIleValAsnValIleuThrAspProAlaPheTyrAsnProGlu 260
Db 393127 TGTCTACTGTGTGATGACCCGCTACGCTGCTTACAGATCTTAATTTCTACAACTCTGCAAG 393186
Qy 261 MetGlnIlyGlyValGlyValAlaValProIlySerMetAlaValAspSerPheLeuTyr 280
Db 393187 CTC---AAAGCTAAGAAA---TCTAAACCTAAGATGACATGAGGGAAGCTTCTTAT 393240
Qy 281 LeuAspAspSerProTyrIleIleuLeuThrIleuValIleAlaTyrGlyIleGly 300
Db 393241 CTGTTAAGATCTCTTATATGCTTCTTCTTACCTCTTCTAATTTGCTTACGGAATCTGC 393300
Qy 301 IleAsnLeuIleGluValIleThrTyrIlySerGlnLeuIlyLeuGlnTyrProAsnMetAsn 320
Db 393301 ATTAACTCTGTGAGTAACTTGAAGAAAGCCAACTCAAAATGCAATCTCTTATCCAAAT 393360
Qy 321 AspTyrSerGlnPheMetGlyAsnPheSerPheTyrIlyValValSerValIle 340
Db 393361 GACTACAGCGCATTTATGGGAACTTCTTCTTCTGACAGAGTTGTATCTGATTTGTGTA 393420
Qy 341 MetLeuPheValGlyGlyAsnValIleArgIlyPheGlyTyrPheLeuThrGlyValAlaVal 360
Db 393421 ATGCTCTTCATCGGTGATGATGATATCGTATGATTCGTTGATTAACGAGCTTGTGTT 393480
Qy 361 ThrProValMetValIleuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn 380
Db 393481 ACACCTATACAGCTTCTCTGTAACAGAGCTGCTCTTCTGCTCTGCTTATCTTCAAGAT 393540
Qy 381 GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValValVal 400
Db 393541 CATCTCATCAGCTTATGTTGCTGCTTGGGAAACAACACATTAATGCTGCGCTGTGTGTC 393600
Qy 401 GlyAlaIleGlnAlaIleuSerIlySerThrIlyTyrAlaLeuPheAspSerThrIly 420
Db 393601 GGACCATTCATAAAATATCTTCCAAATCAATTAATGCTCTTGTGATGCAACTTAA 393660
Qy 421 GlnMetAlaTyrIleProLeuAspGlnIlyGlnIlyValIlyGlyValAlaAlaIleAsp 440
Db 393661 GAGCTGGCTTACATCCCTTGGATCAGAGCAAAAGTTAAAGGGAAGCTGCTATTGAC 393720
Qy 441 ValValAlaAlaArgPheGlyIlySerGlyValAlaLeuIleGlnIlyLeuLeuVal 460
Db 393721 GTTGTGCGCTGATGTTGCTGTAATCTGGGGATCTTGTGATCCAAAGATCTCTGCTC 393780
Qy 461 IleCysGlySerIleGlyAlaMetThrProTyrIleuAlaValIleLeuPheIleIle 480
Db 393781 GTTTCGGAACATCGGGCTATGATCTCAATCTCAATGCTGCTCTTTCGCAATATTC 393840
Qy 481 AlaIleTyrPheValSerAlaThrIlyLeuAsnIlyLeuPheAlaGlnSerAlaLeu 500
Db 393841 ATGCTCTGCTTGAACATCTGCAACTTAACATAAACAATCTTCTGCTGCACTCCGCTCCT 393900
Qy 501 LysGlnIlyGlnValIle-----AlaGlnIly 508
Db 393901 AAAAGAACGAAATTAGNAGAGCTGACAGCAGAGAGAAAGCTTCTTTCGCGGCTTAA 393960
Qy 509 AspSerAlaProAla 513
Db 393961 GAATCCGCTCCTGCT 393975

```

RESULT 5
AB212889 standard; DNA; 1875 BP.

XX AB212889;
XX 21-JAN-2003 (first entry)

```

XX XX Arabidopsis thaliana stress regulated gene SEQ ID NO 694.
DE XX
XX XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX XX
XX XX Arabidopsis thaliana.
XX OS
XX MO200216655-A2.
XX
XX 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US26685.
XX
XX 24-AUG-2000; 2000US-227866P.
XX 26-JAN-2001; 2001US-26647P.
XX 22-JUN-2001; 2001US-300111P.
XX
XX (SCRI ) SCRIPPS RES INST.
XX PA
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Harper JF, Kreps J, Wang X, Zhu T;
XX
XX WPI; 2002-304127/34.
XX
XX Identifying a stress condition to which a plant cell has been exposed
XX PT and producing plants with increased tolerance to these abiotic stresses
XX PT
XX
XX Claim 144; SEQ ID NO 694; 577bp + Sequence Listing; English.
XX
XX The invention relates to identifying a stress condition to which a plant
XX CC cell has been exposed, comprising:
XX CC (a) contacting nucleic acid representative of expressed polynucleotides
XX CC in the plant cell with an array or probes representative of the plant
XX CC cell genome; and
XX CC (b) detecting a profile of expressed polynucleotides in the plant cell
XX CC characteristic of a stress response. The method is useful in the
XX CC production of transgenic plants, cells and seeds and in producing plants
XX CC with increased tolerance to abiotic stress. The present sequence is that
XX CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
XX CC in methods of the invention.
XX CC Note: The sequence data for this patent is not represented in the printed
XX CC specification but is based on sequence information supplied to Derwent by
XX CC the European Patent Office.
XX
XX XX
XX SQ Sequence 1875 BP; 466 A; 416 C; 455 G; 538 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 5,93e-135 Length: 1875
XX Score: 1333.00 Matches: 259
XX Percent Similarity: 69.60% Conservative: 89
XX Best Local Similarity: 51.80% Mismatches: 140
XX Query Match: 50.68% Indels: 12
XX DB: 24 Gaps: 3
XX
XX US-09-869-433-2 (1-515) x AB212889 (1-1875)
Qy 16 LeuTyrProIleHisThrIleGlnLeuIlyValLeuProMetPheLeuMetPhePhe 35
Db 313 ATTTCGCTGTGGAGTTGCACTTGAAAGATTATCCCTTATGAGATTGATGTTCTTT 372
Qy 36 CysIleThrPheAsnTyrThrValIleuArgAspThrIlyAspThrIleuIleValIlyAla 55
Db 373 TGTATTTCTTTCAATTACACAAATTCGAGGATACAAAGATGCTTGTGGTGATCGCGC 432
Qy 56 ProGlySerGlyAlaGluAlaIleProPheIleIlySerPheTyrPheValIleProCysAla 75
Db 433 AAAGGAAGTCTGCTGAGTATTAACCTTTTGAAGACTTGGGATTCCTTCATGAGCC 492
Qy 76 IleIlePheMetLeuIleTyrAlaIlyLeuSerAsnIleuSerIlyGlnAlaLeuPhe 95
Db 493 ATTGCGTTATGCTCTCTACACTTCAAACTCCAAATGTTCTCTCCAAAGAGGCTGTGTT 552

```

```

Qy 96 TyrAlaValGlyThrProPheLeuIlePhePheAlaLeuPheProThrValIleTyrPro 115
Db 553 TACACTGTTATGTCCTTTTCATCATCTTACCTTGGGGCTTTGGTTTGTCATGATACCT 612
Qy 116 LeuArgAspValLeuHisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProPro 135
Db 613 CTCAGCACTATATTCACCCCGAGAGCTCTCGAGATTAAGCTCTTACCAACCTCGGCCCA 672
Qy 136 GlyLeuLeuGlyLeuValAlaIleLeuArgAspThrPheAlaAlaPheTyrValLeu 155
Db 673 AGATTCAAGGGTCTCTTATGCAATATTTGGGATTTGAGATTTCTGTTTGTATATCTTATG 732
Qy 156 AlaGluLeuTyrPGLySerValMetLeuSerLeuMetPheTyrPGLyPheAlaAsnGluIle 175
Db 733 GCTGAGCTTTGGGGAGTGTGTGTCTCAAGTCTCTTCTGGGGCTTTGGTCTATATGATC 792
Qy 176 ThrIysIleHisGluAlaIleAspArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSer 195
Db 793 ACAACTGTGATGAGCCAGCAAGAAATTCATCTTGTTCGAGCTTGAGCCATGTTGCA 852
Qy 196 LeuLeuAlaSerGlyArgAlaIleValTrrpAlaSerIysLeuArgAlaSerValSerGlu 215
Db 853 CTGATTCTCTAGAGAAAGCCGTAAGAACTTCTCTTAACCTTGAGAAAGAAATCTGTGCT 912
Qy 216 GlyValAspProTrrpGlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGly 235
Db 913 GAGATTGACGGCTGGCGACGTTTCGTTGAAGCAAGATGACATTTGTGTGGATGGGA 972
Qy 236 LeuValLeuMetAlaSerTyrTrrpIleAsnIysAsnValLeuThrAspProArgPhe 255
Db 973 CTCGCCATTGTCTCTCTATTTGGTGGGTGATATATGTTCTCTTCAACCGGTATGC 1032
Qy 256 TyrAsnProGluGluMetGlnIlysglyIlysglyAlaIysProIysMetAsnMetIys 275
Db 1033 AAGAACAGAAAGAG-----AAACCGAATAGGAAACGATG 1068
Qy 276 AspSerPheLeuTyrLeuAspArgSerProTrrpIleLeuLeuThrLeuValIle 295
Db 1069 GAAAGCTTGAAGTCTTGTGTATCATCACCATTCATTAAGATCTTGACTTACTGTGTC 1128
Qy 296 AlaTrrpGlyIleCysIleAsnLeuIleGluValThrTrrpIysSerGlnLeuIysLeuGln 315
Db 1129 GCATACGGTATTAATCATCTTGTGTGAAGTCACATGGAATCAAAGCTTAAAGCTCAG 1188
Qy 316 TyrProAsnMetAsnAspTyrSerGluPheMetGlyAsnPheSerPheTrrpIleVal 335
Db 1189 TTCCTTACCCGAAATGAGTACTCAGCATTTATGGAGACTTCTCAACCTGACGGGTCTT 1248
Qy 336 ValSerValLeuIleMetLeuPheValGlyIysAsnValIleArgIysPheGlyTrrpLeu 355
Db 1249 GCACATTCACAAATGATGCTT---CTCAGCCAATTCGATTCAATAGTATGTTGGGGA 1305
Qy 356 ThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleValaPhePheAlaLeu 375
Db 1306 GTAGCTGGAAGATCACCCCAAGCTGTCTGCTATTAAGTGTGGTGGCTTCTTCTCTCTA 1365
Qy 376 ValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMet 395
Db 1366 ATATGTTTGGCGGCCCATTCGACCACTTGTTCGCAAGCTTGATGACACCGGTACTT 1425
Qy 396 LeuAlaValValAlaGlyAlaIleGlnAsnIleLeuSerIysSerThrIysTyrAlaLeu 415
Db 1426 GCACCTGTGTATGTCGGTCCCTTCAGATATCTTCAGCAAGATGCGCAAGTACAGCTTG 1485
Qy 416 PheAspSerThrIysGluMetAlaTrrpIleProLeuAspGlnGluIysValIysGly 435
Db 1486 TTCACCTTCGAAAGAAATGGCTATATCCATTCATGATGAGACACCAAGATTAAAGGC 1545
Qy 436 LysAlaIleAlaIleAspValValAlaAlaArgPheGlyIysSerGlyValAlaLeuIleGln 455
Db 1546 AAACCTGGATTTGACGTGTGTGCAACCATTAAGGAAATAGGGGAGCTTTTATATACG 1605
Qy 456 GlnGlyLeuLeuValIleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValIle 475

```

```

Db 1606 CAGTTCATGATCTTATCTTGGATCACTACAGGAATTCAGAGCGGTATCTTGAATATC 1665
Qy 476 LeuLeuPheIleIleAlaIleTrrpLeuValSerAlaThrIysLeuAsnIysLeuPheLeu 495
Db 1666 TTGTTGTTATTTGTCATCGCTGTGTAGCTCAGCTTAAGTGTGCTGGAGGACAGTTTC--- 1722
Qy 496 AlaGlnSerAlaLeuIysGluGlnGluValAlaGlnGluAspSerAlaProAlaSerSer 515
Db 1723 -----AACACCTTGGGTCTGAAAGAGAGCTTGAAGAGAAATGAGACAGAGCTTCATCG 1776

RESULT 6
AAC36046
ID AAC36046 standard; DNA; 2146 BP.
AC AAC36046;
DE 17-OCT-2000 (first entry)
AC Arabidopsis thaliana DNA fragment SEQ ID NO: 12350.
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
EN EP1033405-A2.
PD 06-SEP-2000.
PE 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 07-MAY-1999; 99US-0132486.
XX 11-MAY-1999; 99US-0132487.
XX 14-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.

```

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144336.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0151065.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158237.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159229.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 1,35e-134
Score: 1330.50
Percent Similarity: 68.88%
Best Local Similarity: 50.88%

Length: 2146
Matches: 260
Conservative: 92
Mismatches: 136

This invention describes a novel transformed plant (A), and its descendants, in which the regulatory sequences and/or the gene copy number of an ATP/ADP (adenosine tri-/di-phosphate) translocator gene (1) has been altered so that, compared with the wild-type plant, it has a different content of one or more amino acids (aa). The invention also describes (1) (1), for use in plants, having an approximately 1.7 kb sequence (S1), given in the specification, from Arabidopsis thaliana, EMBL Z49227; (2) a genetic construct (GC) containing the (1) and operably linked regulatory sequences; (3) a vector containing the (1) or GC; and (4) seeds, tissues, cells and replicative materials of (A). (A) are useful for human or animal feeding, also (including their cells, tissues and extracts) in agriculture, the animal feed and pharmaceutical industries, and the health service. Altering the activity of (1) increases the amount of ATP in chloroplasts and the amount of energy of available for biosynthesis. Specifically (A) have increased contents of one or more essential aa, so are of greater nutritional value than wild-type plants.

Sequence 1896 BP; 473 A; 369 C; 445 G; 609 T; 0 other;

Alignment Scores:

Pred. No.: 5.3e-132 Length: 1896
Score: 1306.00 Matches: 259
Percent Similarity: 69.34% Conservative: 87
Best Local Similarity: 51.90% Mismatches: 141
Query Match: 49.66% Indels: 13
DB: 22 Gaps: 3

US-09-869-433-2 (1-515) x AAF61374 (1-1896)

Qy 16 LeuTPProIleHisThrHisGluLeuValLeuProMetPheLeuMetPhePhe 35
Db 308 TTAATGGG-ATTGAACCTGTGACCTTAAGAAATATTATACCACTGGGCGCATGTTCTT 366
Qy 36 CysIleThrPheAspTyrThrValLeuArgAspThrIleAspThrLeuIleValGlyAla 55
Db 367 TGTATTCGTGTTAATTATACAACTCTTAGGATACCTAGATGCTGTGTAACACT 426
Qy 56 ProGlySerGlyAlaGluAlaIleProPheIleuSerPheTyrLeuValValProCysAla 75
Db 427 AAGAGGTCACGTGTCGATGATTAATCCCTTCTTGAACCTGGGTAATTCCTATGCT 486
Qy 76 IleIlePheMetLeuIleTyrAlaIleuSerAsnIleLeuSerIleGlnAlaLeuPhe 95
Db 487 ATTGATTCATGCTTTTGTACACAAAGTGGCTAATGCTGTCTCAAGAGGCTCTTTT 546
Qy 96 TyrAlaValGlyThrProPheLeuIlePhePheAlaLeuPheProThrValIleTyrPro 115
Db 547 TATACGTGTTATACCTCTTTATGTCATTCCTTGGGCGCTTGTGTTTGTATCTCT 606
Qy 116 LeuArgAspValLeuHisProThrGluPheAlaAspArgGlnAlaIleLeuProPro 135
Db 607 CTTAGCAATTACTTACCCCTACAGCTTTGCTGTAAGCTTCTCAATACCTTGGTCA 666
Qy 136 GlyLeuLeuGlyLeuValAlaIleLeuArgAsnTyrThrPheAlaIlePheTyrValLeu 155
Db 667 AGATTCTTGAACCAATGCTATTCTGAGATCTGAGATTCCTGCTTCTATGCTATG 726
Qy 156 AlaGluLeuTyrGlySerValMetLeuSerLeuMetPheTyrGlyPheAlaAsnGluIle 175
Db 727 GCTGAGCTTGGGGAAGGTGGTGGTTCAGTACTCTTTGGGGAATTCATACAGATTC 786
Qy 176 ThrIleIleHisGluAlaIleArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSer 195
Db 787 ACGAGCTGTCATGAGGCTTAAGATTCATCTTGTGTGACCTTGGAGCAAGTGTGCT 846
Qy 196 LeuLeuAlaSerGlyArgAlaIleValTyrAlaSerIleuArgAlaSerValSerGlu 215
Db 847 CTATATTTCTCTGTCGACACAGTAAAGTATTTCTTACTTGAAGAGCTCTTTAGTCTCT 906
Qy 216 GlyValAspProTyrGlyIleSerLeuArgLeuMetAlaMetThrIleValSerGly 235
Db 907 GAGGTGATGATGGTGGCTATCTCTCTGAAGAAAGATGATGATGATGATGATGATG 966

Qy 236 LeuValLeuMetAlaSerTyrTyrTrpIleAsnIleValLeuThrAspProArgPhe 255
Db 967 GGGGCAATCTGTTCTTTTACTGATGGTGAATGAATGTGCTCTCCCAACTCTG--- 1023
Qy 256 TyrAsnProGluGluMetGlnIleGlyValIleValProIleuSerMetAsnMetIle 275
Db 1024 -----ACGAG 1062
Qy 276 AspSerPheLeuTyrLeuAspArgSerProTyrIleLeuLeuThrLeuValIle 295
Db 1063 GAGAGCTTGAAGTCTTGGTCTCTTCAAAATATATCAAGGATCTTGCACATTTGGTGT 1122
Qy 296 AlaTyrGlyIleCysIleAsnIleGluValIleThrTrpIleSerGlnIleValLeuGln 315
Db 1123 GCATATGCACTTACTATCAACCTTGTGAAGTTTCAAGAGATCAAGAGTCAAGCTCAAG 1182
Qy 316 TyrProAsnMetAsnAspTyrSerGluPheMetGlyAsnPheSerPheTyrTrpIleVal 335
Db 1183 TTCCCAAGCCCAATGAATATCTCTCATTCATGGGTGATCTTCACTGCTACTGAGATA 1242
Qy 336 ValSerValLeuIleMetLeuPheValGlyValAsnValIleArgIlePheGlyTyrLeu 355
Db 1243 GCAACTTTCACAAATGATGTG---TTAAGTCATGATTTTCGACAAAGTATGGGTGGGA 1299
Qy 356 ThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeu 375
Db 1300 GCAGACGCCAAGATTAACACTTACAGTCTTGTCTTACCGAGTGTGTTCTTCTCCCTG 1359
Qy 376 ValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMet 395
Db 1360 CTTTGTGTTGGGCGACCTCTAGCACTTACTCTCGAAGTTTGAGATGATGATCTCTCTTA 1419
Qy 396 LeuAlaValValGlyAlaIleGlnAsnIleLeuSerIleSerThrIleTyrAlaLeu 415
Db 1420 GCACCTGTCAATGGGTGCAATGCAACATTTTCAGTAAAGTGCAGAAAGTATAGTTTG 1479
Qy 416 PheAspSerThrIleGluMetAlaTyrIleProLeuAspGlnIleGlnIleValIleGly 435
Db 1480 TTTTACCCCTGCAGAAAGATGCTCTTACATTCCTTTGATGAGACACCAAGTTAAAGCG 1539
Qy 436 LysAlaAlaIleAspValValAlaAlaArgPheGlyIleSerGlyIleAlaLeuIleGln 455
Db 1540 AAGGCAACATTCATGATGTTGCTGCATTCACCTGGAGAAAGCTTGAGAGCTTATCAAA 1599
Qy 456 GlnGlyLeuLeuValIleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValIle 475
Db 1600 CAGTTCAATGATTTGACCTTGTGCTTGCCTGCGACGCTGCACACCTTACCTTGGCGGTG 1659
Qy 476 LeuLeuPheIleIleAlaIleTyrLeuValSerAlaThrIleValLeuValLeu 495
Db 1660 CTCTTAGTAATTTGTTCTGATGATGTTGGAGACACCAAGCTTGTGATGACAGCTTC--- 1716
Qy 496 AlaGlnSerAlaLeuIleGluGlnIleValAlaGlnIleuSerAlaProAlaSer 514
Db 1717 -----ACTCAATTAACCAAGAAAGATCTTGAAGAAAGATGAGAGAGCATCG 1767

RESULT 8
AAF61373 standard; cDNA; 1770 BP.
ID AAF61373;
AC AAF61373;
DT 25-MAY-2001 (first entry)
XX A. thaliana chloroplast ATP/ADP translocator cDNA.
XX ATP/ADP translocator; chloroplast; plant; adenosine triphosphate;
XX adenosine diphosphate; animal feed; energy; biosynthesis;
XX nutritional value; ss.
XX Arabidopsis thaliana.
XX OS

QY 491 enlyseuphenuaiaaginserralaLeuylglnuVala1a 506
 DB 1715 ACAGCTTGCGCTCTGAAGAGAGCTTGAG-AGGAAATGAGAGAGCT 1760

RESULT 9
 AAF84486
 ID AAF84486 standard; DNA; 1823 BP.
 XX AAF84486;
 AC AAF84486;
 DT 02-JUL-2001 (first entry)
 XX
 DE Chlamydia pneumoniae Npt2cp (ADP/ATP translocase) gene.
 XX
 KW Npt2cp; ATP/ADP translocase; respiratory tract infection; pneumonia;
 KW bronchitis; sinusitis; atherosclerosis; asthma; vaccine; antimicrobial;
 KW antibacterial; diagnosis; ds.
 XX
 OS Chlamydia pneumoniae.
 XX
 FH Key Location/Qualifiers
 FT CDS 101..1723
 FT /*tag= a
 FT /product= "Chlamydia pneumoniae Npt2cp protein"
 FT /function= "ADP/ATP translocase"
 XX
 FN WO200121803-A1.
 XX
 PD 29-MAR-2001.
 XX
 PF 15-SEP-2000; 2000WO-CA01087.
 XX
 PR 17-SEP-1999; 99US-0154326.
 XX
 PA (AVENT) AVENTIS PASTEUR LTD.
 XX
 PI Murdin AD, Oomen RP, Wang J, Dunn P;
 XX
 DR WPI, 2001-316102/33.
 DR P-PSDB; AAB73452.
 XX
 PR New Npt2cp (ADP/ATP translocase) polypeptides and nucleic acids
 PT encoding the polypeptides useful for treating, preventing or diagnosing
 PT Chlamydia infections, particularly infections caused by Chlamydia
 PT pneumoniae -
 XX
 PS Claim 2; Fig 1; 79pp; English.
 XX
 CC This sequence represents the Chlamydia pneumoniae Npt2cp (ADP/ATP
 CC translocase) gene. Chlamydia pneumoniae Npt2cp nucleotides and
 CC proteins may be used for preventing, treating or diagnosing
 CC Chlamydia infections, particularly those caused by Chlamydia
 CC pneumoniae. Chlamydia pneumoniae can be responsible for both upper and
 CC lower respiratory tract infections. It is the third most common cause of
 CC community acquired pneumonia, and can also cause bronchitis and
 CC sinusitis. It is also linked to diseases other than respiratory tract
 CC infections, being associated with atherosclerosis and asthma.
 CC Npt2cp nucleotides may be used for the recombinant production of the
 CC protein, as genetic vaccines, in the construction of vaccine vectors
 CC such as poxviruses, and in the construction of attenuated Chlamydia
 CC strains which can over-express an Npt2cp nucleotide, or which can
 CC express it in a non-toxic, mutated form. Npt2cp polypeptides can be used
 CC in vaccine compositions, and may also be used as diagnostic reagents for
 CC detecting the presence of anti-Chlamydia antibodies in a sample.
 CC
 XX
 SQ Sequence 1823 BP; 450 A; 359 C; 396 G; 618 T; 0 other;

Alignment Scores:
 Pred. No.: 1,93e-101 Length: 1823
 Score: 1025.50 Matches: 214
 Percent Similarity: 59.12% Conservative: 107
 Best Local Similarity: 39.41% Mismatches: 185

Query Match: 38.99% Indels: 37
 DB: 22 Gaps: 9
 US-09-869-433-2 (1-515) x AAF84486 (1-1823)

QY 1 MetThrlysrhrglugluysrProphedlylsleuArgSerPheLeuTrrpIleHis 20
 DB 101 ATGCAGTCATCAGAAAGTGAACCCCTTTTCAAGGCTGGCGGACATCTTTGTCTATTTAT 160
 QY 21 ThrHsgluLeuylslyValLeuProwePheLeuMetPhePheCysIleThrPheAsn 40
 DB 161 AAATCAGAAATTTCTAAGTTTGTTCCACTATTTCTACTAGGCTTTTTCGTTGCTTTAAC 220
 QY 41 TyrThrValLeuArgAspThrlyAspThrlyLeuIleValIleAlaProGlySerGlyVala 60
 DB 221 TACTGCCCTGCTGAAAACATGAAGATATCTGTGCTATTCGCTGATCGCTTCAGATCGTGGGCA 280
 QY 61 GluAlaIleProPheIlelysrPheTrpLeuValIleProCysAlaIleIlePheMetLeu 80
 DB 281 GAAGTGAATTCCTTCCTTAAGCTTTGGGAAATGTCCCGGAGCTGTTATTGTTACTATG 340
 QY 81 IleTyrAlaIlysrSerAsnIleLeuSerlysglnAlaLeuPheTyrAlaValGlyThr 100
 DB 341 GCTATGGGTGGTGAAGCAGCGATCTTCGCGATACCGTTTATTGCTTCATGCGC 400
 QY 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
 DB 401 GCATTCCTGTTGTTTCTTCTCTGTTGCTGTGATCATTTTCTGTGAGGGGATAGCTTG 460
 QY 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140
 DB 461 CATCTCAACTCTCTCGCTGATTAATTAACAAGCTCTTCCCTCAAGAGCTTCGTGGTTT 520
 QY 141 ValAlaIleLeuArgAsnTrrPheAlaIlePheTyrValIleAlaGluLeuTrrGly 160
 DB 521 ATTGTGATGTCCTTACTGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 580
 QY 161 SerValMetLeuSerLeuMetPheTrrPglyPheAlaAsnGluIleThrlyIleIleGlu 180
 DB 581 TCGGTTGCTCTTTCGATGCTTCTTCGGGACTACCATCATCATCATCATCATCATCATCAT 640
 QY 181 AlaIlysrPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
 DB 641 GCGGCGCGTTTTCAGCTTATCATATACAGATTAATTAATTAATTAATTAATTAATTAATTA 700
 QY 201 ArgAlaIleValIleTrrPalaSerlyLeuArg---AlaSerValSerGluGlyValAspPro 219
 DB 701 GAAATCTCTTAATGATGGGAAACAAATTTGTCCTACTCTTTCATGATGATTC 760
 QY 220 Trp---GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValIleSerGlyLeuValLeu 238
 DB 761 TGGCACTCTGTAATGCTCACTTACCATGCTGATCACT---TGTTCGTTTAATTAATG 817
 QY 239 MetAlaSerTyrTrrPrrIleAsnIlysrValIle-----LeuThrAspProArgPheTyr 256
 DB 818 ATC-----TGGCTATATAGGGGATTCATCATTTGACTATGATGATCTTCGATC 865
 QY 257 AsnProGluGluMetGlnIlysrGlyIlysrGlyVala----- 268
 DB 866 CCTTCATCTAGACCTGTCTTGGCAGAAAGAGGACACTACTGCTAACTGAAGAAAG 925
 QY 269 ---LysProIlysrMetAsnMetIlysrAspSerPheLeuTyrLeuAspArgSerProTrrIle 287
 DB 926 AAAAACCCTTAAGCAAGCAAGCTAGAAACCTTTTCTTACACTCATTCAGTCTCGTATTTTA 985
 QY 288 LeuLeuLeuThrLeuLeuValIleAlaTrrGlyIleCysIleAsnLeuIleGluValIleThr 307
 DB 986 TTAGGCTCGCTATATATGTCCTATTAATTAATTTGGATGATCATTAATGGAAGTCGT 1045
 QY 308 TrpIlysrGlnLeuIlysrLeuGlnTrrProAsnMetAsnAspIlysrSerGluPheMetGly 327
 DB 1046 TGGAAAGATCAAGTTTACCGAGATTTTACAGTTTCACGATGAATTAATTAATTAATTAATGAGT 1105

QY 328 AsnPhSerPheThrGlyValValSerValLeuLeuMetLeuPheValGlyValAsn 347
 Db 1106 AGATCACTACCCCTATTGGGCTGCTTCTATTTAGACAGTACCTTACCGGACAG 1165
 QY 348 ValIleArgLysPheGlyThrPheLeuThrGlyValLeuValThrProValMetValLeuLeu 367
 Db 1166 TGTATCCGTAAATGGATGAGTGTGCTGCTTATGACATTCATTGTATGTTAGTT 1225
 QY 368 ThrGlyLeuValPhePheLeuValIle---PheArgAsnGlnIleSerGlyLeuVal 386
 Db 1226 TCAGAGCTGCTCTTTTTCGAACTATTTTTCGCAAAAGACATCTTATTTTGGG 1285
 QY 387 AlawerPheGlyThrThrProLeuMetLeuValValValGlyValIleGlnAsnIle 406
 Db 1286 GGAGTCTTGGAAAGACCTCTGCTTACCTGCTGAGCTGAGGAGGATGCAAAATGTC 1345
 QY 407 LeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLysGluMetAlaTyrIlePro 426
 Db 1346 CTATCCCGGGGAGCAAAATTACGTTCTTGATCAACCAAGAAATGGCCTTATCCCA 1405
 QY 427 LeuAspGlnGluGlnLysValLysGlyValValAlaIleAspValValAlaIleArgPhe 446
 Db 1406 CTTTCTCCAGAGATTAATAATCATGAGAAACCGCATGTGATGGTTCGTTCAAGATA 1465
 QY 447 GlyLysSerGlyValAlaLeuIleGlnGlnLysLeuValIleCysGlySerIleGly 466
 Db 1466 GGAAAGCTGAGGCTCTTATTTATACCAAGGCTGCTGTTATTTCTCTCTGTTGCA 1565
 QY 467 AlawerThrProTyrLeuAlaValIleLeuPheIleIleAlaIleThrPheValSer 486
 Db 1526 GCAAGTTTAAACGATCGCCCTGCTTCTCATTTATTAATGATGCGTTGATGGCGTT 1585
 QY 487 AlaThrLysLeuValLysPheLeuAlaGlnSerIle-----LeuLys 501
 Db 1586 GTTGCTATATCGGTAAAGAAATCACTACTAGACTCTGATGCTGAGCAACCTTAAA 1645
 QY 502 -----GluGlnGluValAlaGlnGlu 508
 Db 1646 CAACCTAAAGACCTTCTCTTCAATGATGATGATGACCCAGGATCTGTAAGAAAGAA 1705
 QY 509 AspSerAla 511
 Db 1706 GAATGCT 1714
 Db 1706 GAATGCT 1714
 RESULT 10
 AAX91990/c
 ID AAX91990 standard; DNA, 1230025 BP.
 AC AAX91990;
 XX
 DT 13-SEP-1999 (first entry)
 XX
 DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
 XX
 KM Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KM vaccine; neutralising epitope; ss.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN MO9927105-A2.
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-IB01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GENST) GENSET.
 PI Griffiths R;
 XX

DR WPI, 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 XX
 PS Claim 1; Page 291-611; 1912pp; English.
 XX
 CC The present sequence represents the complete genome of Chlamydia
 CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes
 CC respiratory disease such as pneumonia and bronchitis and is thought
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
 CC AAY34584-Y35879 can be used in immunogenic compositions as vaccines.
 CC Vectors containing C. pneumoniae nucleotide sequences can also be
 CC used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae.
 XX
 SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;
 Alignment Scores:
 Pred. No.: 2.65e-97 Length: 1230025
 Score: 1025.50 Matches: 214
 Percent Similarity: 59.12% Conservative: 107
 Best Local Similarity: 39.41% Mismatches: 185
 Query Match: 38.99% Indels: 37
 DB: 20 Gaps: 9
 US-09-869-433-2 (1-515) x AAX91990 (1-1230025)
 QY 1 MetThrLysThrGluGlnLysPheGlyLysLeuArgSerPheLeuThrProIleHis 20
 Db 717505 ATGCAGTCATCAGAAAGGAAACCTTTTCAAGGCTGGGGCATATCTTGTCCATTTAT 717446
 QY 21 ThrHisGluLeuLysValLysPheLeuPheLeuMetPheCysIleThrPheAsn 40
 Db 717445 AAATCAGAAATTTCTAAAGTTTGTTCACATTTCTACTAGCGTTTTCGTTGCTTAAAC 717386
 QY 41 TyrThrValLeuArgAspThrLysAspThrLeuIleValGlyAlaProGlySerGlyAla 60
 Db 717385 TACTGCTGCTGAAAAACATGAAGATACCTGCTGATTCGTTTCAGATGCTGGGCA 717326
 QY 61 GluAlaIleProPheIleLysPheThrPheValValProCysAlaIleIlePheMetLeu 80
 Db 717325 GAAGTATTCCTCTTCTTAAGGTTTGGGAAATGTCCCGGAGCTGTATTTGTTACTATG 717266
 QY 81 IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr 100
 Db 717265 GTCTATGCGGTGTTAGGCACTGCTATCCCGGATACGTTTATTTGCTTCATGGCC 717206
 QY 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
 Db 717205 GCATTCCTTGTTTCTTCTTCTGCTGTTGCTGATCATTTATCTGATGGGATAGCCTG 717146
 QY 121 HisProThrGluPheAlaAspArgLysGlnAlaIleLeuProGlyLeuLeuGlyLeu 140
 Db 717145 CATCTCAACTCTCTCGCTGATAATTAATTAACAGAGCTCTTCTCAAGAGCTTCGGTITT 717086
 QY 141 ValAlaIleLeuArgAsnThrPheAlaAlaPheTyrValLeuAlaGluLeuTyrGly 160
 Db 717085 ATTGTATGTCGCTGCTACTGAGATTAAGATTTATTAATGATGATGAGCTGTGAGT 717026
 QY 161 SerValMetLeuSerLeuMetPheTyrGlyPheAlaAsnGluIleThrLysIleHisGlu 180
 Db 717025 TCGTTGTTCTTTCGATGTTGTTCTGGGAGCTGACCAATGATTTCTCAAAATTAATCGAA 716966
 QY 181 AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuValAspSerGly 200
 Db 716965 GCGGGCGTTTATTCGCTTATCAATACAGGATTAATTAATTTCTTCCATATGCGGAGGA 716906
 QY 201 ArgAlaIleValIleTyrAlaSerLysLeuArg---AlaSerValSerGluValAspPro 219
 Db 716905 GAATCTCTATTTGATGGGAAACAAATTTTGTGCTTACCTTGTGATGATGATTC 716846

| | | | |
|----|--------|---|--------|
| Oy | 220 | Trp---GlyIleSerLeuArgLeuMetAlaMetThrIleValSerGlyLeuValLeu | 238 |
| | | | |
| Db | 716845 | TGGCACTCTGTATATGCTCAACTTGGACCAATGCGATCACT---TGTTCTGGTTTAATTAATG | 716789 |
| Oy | 239 | MetAlaSerTyrTrpTrpIleAsnIysAsnVal-----LeuThrAspProArgPheTyr | 256 |
| | | | |
| Db | 716788 | ATC-----TGGCTAATATAGGCGGATCATCATTTGACTATTGATCTACTTGATCTTGCAATC | 716741 |
| Oy | 257 | AsnProGluGluMetGlnIlySgIlyValSgIlyValA----- | 268 |
| | | | |
| Db | 716740 | CCTCATCTAGACGTGCTCTGGCAGAGAGGGGAGCGCTACTGCTAATCTAAAGAAAAG | 716681 |
| Oy | 269 | ---LysProIysMetAsnMetLysAspSerPheLeuTyrLeuAspArgSerProTyrIle | 287 |
| | | | |
| Db | 716680 | AAAAAACCTTAAAGCCAAAGCTGAAACCTTTTCTTACACCTCACTTACAGTCTCGTTATTTA | 716621 |
| Oy | 288 | LeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCysIleAsnLeuIleGluValThr | 307 |
| | | | |
| Db | 716620 | TTAGGCGCTCGTATATTATGTCCTAATCTTAATATTGGTATCATCATTTGCAAGTCGTT | 716561 |
| Oy | 308 | TrpIlySerGlnLeuLeuValLeuGlnTyrProAsnMetAsnAspTyrSerGluPheMetGly | 327 |
| | | | |
| Db | 716560 | TGGAAAGATCAAGATTAGCCAGATTATACAGTCTTCACGTGAATTCAATGGGTATATAGAT | 716501 |
| Oy | 328 | AsnPheSerPheTrpThrGlyValIleSerValIleIleMetLeuPheValGlyIlyAsn | 347 |
| | | | |
| Db | 716500 | AGATACATCAACCTCATATTGGCGTCGTTGTGATTTAGCAGCTGTACTCTTACCGAGACAG | 716441 |
| Oy | 348 | ValIleArgIlyPheGlyTrpLeuThrGlyAlaLeuValThrProValMetValLeuLeu | 367 |
| | | | |
| Db | 716440 | TGTATCCGTAATGGGAGATGAGACTGCGGTCGTTTGTGATCTCATCTCAATTGGTAATGTTAGTT | 716381 |
| Oy | 368 | ThrGlyIleValPhePheAlaLeuValIle---PheArgAsnGlnAlaSerGlyLeuVal | 386 |
| | | | |
| Db | 716380 | TCAGAGACGCTCTTTTTCGGAACATATTGTCGCAAAAAGAGACATCTCTAATTTTGGG | 716321 |
| Oy | 387 | AlaMetPheGlyThrThrProLeuMetLeuAlaValValGlyAlaIleGlnAsnIle | 406 |
| | | | |
| Db | 716320 | GGAGTCTTGGAAATGACACCTCTGGCTCTAGCTGCTGACCTGGACCTGGAGGGATGCAAAATGTC | 716261 |
| Oy | 407 | LeuSerLysSerThrIleTyrAlaLeuPheAspSerThrIlySgIleMetAlaTyrIlePro | 426 |
| | | | |
| Db | 716260 | CTATCCCGGGGACGAAATTTACGTTCTTGTATCAACCAAGGAATATGGCTTTATCCCA | 716201 |
| Oy | 427 | LeuAspGlnGluGlnIlyValIlySgIlyValAlaIleAspValValAlaAlaArgPhe | 446 |
| | | | |
| Db | 716200 | CTTCTCCAGAGATAAATAATCATGGGAACCGCGCATGTGATGTCGTTCAAGATATA | 716141 |
| Oy | 447 | GlyIlySerSerGlyValAlaLeuIleGlnGlnIlyLeuLeuValIleCysSgIleSerIleGly | 466 |
| | | | |
| Db | 716140 | GGAAAGCTCGAGGCGCTCTTAAATTTACCAAGGCGCTGTTATTTCTCTCTGTGCA | 716081 |
| Oy | 467 | AlaMetThrProTyrLeuAlaValIleLeuLeuPheIleIleAlaIleTrpLeuValSer | 486 |
| | | | |
| Db | 716080 | GCAAGTTAAACGTATCGCCCTTAAGTCTTCTCATTTATATGTCGTTTGATGCGGTT | 716021 |
| Oy | 487 | AlaThrIlyLeuAsnIlyLeuPheLeuAlaGlnSerAla-----LeuIys | 501 |
| | | | |
| Db | 716020 | GTTGCGTATATCGTAAAGAAATACACTCTAGAGCTGCTGATGCTGTAGCAACCTTGAAA | 715961 |
| Oy | 502 | -----GluGlnGluValAlaGlnGlu | 508 |
| | | | |
| Db | 715960 | CAACCTTAAGACCTTCTCTTCAATCGTACGTAGAGCCACAGAAATCTGTAGAACAAAGA | 715901 |
| Oy | 509 | AspSerAla | 511 |
| | | | |
| Db | 715900 | GAAATGGCT | 715892 |

RESULT 11

ABL74862

ID ABL74862 standard; cDNA; 264 bp.

CC ABL74862;

XX DT 14-MAY-2002 (first entry)
XX XX
XX DE Corn tassell-derived polynucleotide (cdps) SEQ ID NO:4236.
XX XX
XX KW Corn; corn tassell-derived polynucleotide; cdps; hybrid breeding; CDPs;
KW inheritance; characteristic; growth; development; disease resistance;
KW environmental adaptability; quality; yield; molecular marker;
XX multigene trait; plant breeding; corn tassell; gene; ss.
XX Zea mays.
XX OS
XX PN US2001051335-A1.
XX PD 13-DEC-2001.
XX PF 16-APR-1999; 99US-0294093.
XX PR 21-APR-1998; 98US-082567P.
XX PA (LALG/) LALGUDI R V.
PA (ITOL/) ITO L Y.
PA (SHER/) SHERMAN B K.
XX PI
PI Lalgudi RV, Ito LY, Sherman BK;
XX WPI; 2002-163647/21.
DR
PT Novel purified corn tassell-derived polynucleotide useful for
PT determining altered gene expression, to recover regulatory elements and
PT to follow inheritance of desirable characteristics through hybrid
PT breeding programs -
PS
PS Claim 1; SEQ ID 4236; 201pp; English.
XX
XX The present sequence describes a purified corn tassell-derived
CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
CC selected from those given in ABL70627 to ABL76833. The cdps sequences
CC encode corn tassell-derived polypeptides (CDPs). The cdps sequences (1)
CC can be used for determining altered gene expression, to recover
CC regulatory elements and to follow inheritance of desirable
CC characteristics through hybrid breeding programs. (I) are also useful
CC in the evaluation, and alteration of desired characteristics associated
CC with growth and development, disease resistance, environmental
CC adaptability, quality and yield, and as molecular markers for studying
CC inheritance of multigene traits in a plant breeding program. (I) can be
CC used to produce a tassell-specific profile of gene transcription, a
CC transcript image, to clone regulatory elements for use in transformation
CC vectors, to express a polypeptide, to identify, isolate or extend
CC identical or related corn tassell nucleic acid sequences from DNA
CC libraries, in nucleic acid hybridisation or amplification technologies,
CC as query sequences to determine homology of known sequences, as probe
CC for use in Southern or Northern hybridisation, and to identify the
CC presence of and/or to determine the degree of similarity between two
CC (or more) nucleic acid sequences.
XX
SQ Sequence 284 BP; 66 A; 55 C; 75 G; 86 T; 2 other;

Alignment Scores:

| Pred. No.: | Length: |
|-------------------------------|------------------|
| Score: 1,03e-18 | 284 |
| Percent Similarity: 256.00 | Matches: 56 |
| Best Local Similarity: 72.34% | Conservative: 12 |
| Query Match: 59.57% | Mismatches: 26 |
| DB: 9.73% | Indels: 1 |
| | Gaps: 0 |

US-09-869-433-2 (1-515) x ABL74862 (1-284)

377 ILPEPhARgASnGLIALAsercLYLeuVALAlawetPheGIYThrThrProleMetLeu 396
Db 3 TTTCTTGGTGACCACTATGACTCCTTTATAGCAAAGTTGGAGTAGACGCTTTGCTTCG 62
397 ALAvalAlvalAlglYAlAlleGlAnsnlleUenSerlysserThrySTyralaleupe 416

Db 63 GCAGCTCATGTTGGAGCATGACATTTTCAGTAAGATGCAAAATACAGTCTGTT 122
 Qy 417 AsperThirlysglMetAlaTyrlleProleuaspnglulglnlyValleuGlylys 436
 Db 123 GATCTTCGNAAGATGCGATACATTCCTTTGGATGAGATGATGAAAGTMAA 182
 Qy 437 AAlaAlaIleaspValIvalAlaAlaArpHeGlylysserGlyValAlaIleuIleGln 456
 Db 183 GCGGCTATTGATGTTGTGTGCAACCCCTTGAGGAATGTGAGGGTCTGATCCAGCG 242
 Qy 457 GlyleuIleuValIleCyGlySerIleGlyAlaMetThrPro 470
 Db 243 TTCATGATCCT-GTTCATGCTCTCTCGCGAANTGCAACCG 283
 RESULT 12
 ABL76356
 ID ABL76356 standard; cDNA; 246 BP.
 XX
 AC ABL76356;
 XX
 DT 14-MAY-2002 (first entry)
 XX
 DE Corn tassal-derived polynucleotide (cdps) SEQ ID NO:5730.
 XX
 KM Corn; corn tassal-derived polynucleotide; cdps; hybrid breeding; CDPS;
 KM inheritance; characteristic; growth; development; disease resistance;
 KM environmental adaptability; quality; yield; molecular marker;
 KM multigene trait; plant breeding; corn tassal; gene; ss.
 XX
 OS Zea mays.
 XX
 PN US2001051335-A1.
 XX
 PD 13-DEC-2001.
 XX
 PF 16-APR-1999; 99US-0294093.
 XX
 PR 21-APR-1998; 98US-082567P.
 XX
 PA (LALG/) LALGUDI R V.
 PA (ITOL/) ITO L Y.
 PA (SHER/) SHERMAN B K.
 XX
 PI Lalgudi RV; Ito LY, Sherman BK;
 DR WPI; 2002-163647/21.
 XX
 PT Novel purified corn tassal-derived polynucleotide useful for
 PT determining altered gene expression, to recover regulatory elements and
 PT to follow inheritance of desirable characteristics through hybrid
 PT breeding programs -
 XX
 PS Claim 1; SEQ ID 5730; 201pp; English.
 XX
 CC The present sequence describes a purified corn tassal-derived
 CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
 CC selected from those given in ABL70627 to ABL76833. The cdps sequences
 CC encode corn tassal-derived polypeptides (CDPS). The cdps sequences (1)
 CC can be used for determining altered gene expression, to recover
 CC regulatory elements and to follow inheritance of desirable
 CC characteristics through hybrid breeding programs. (1) are also useful
 CC in the evaluation, and alteration of desired characteristics associated
 CC with growth and development, disease resistance, environmental
 CC adaptability, quality and yield, and as molecular markers for studying
 CC inheritance of multigene traits in a plant breeding program. (1) can be
 CC used to produce a tassal-specific profile of gene transcription, a
 CC transcript image, to clone regulatory elements for use in transformation
 CC vectors, to express a polypeptide, to identify, isolate or extend
 CC identical or related corn tassal nucleic acid sequences from DNA
 CC libraries, in nucleic acid hybridisation or amplification technologies,
 CC as query sequences to determine homology of known sequences, as probe
 CC for use in Southern or Northern hybridisation, and to identify the

CC presence of and/or to determine the degree of similarity between two
 CC (or more) nucleic acid sequences.
 XX
 SQ Sequence 246 BP; 58 A; 49 C; 63 G; 76 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,81e-17 Length: 246
 Score: 242.00 Matches: 45
 Percent Similarity: 74.68% Conservative: 14
 Best Local Similarity: 56.96% Mismatches: 20
 Query Match: 9.20% Indels: 0
 DB: 24 Gaps: 0
 US-09-869-433-2 (1-515) x ABL76356 (1-246)
 Qy 348 ValIleArglyspHeGlyThrProleuMetIleuValIleValIleGlnAsnIleu 367
 Db 6 ATCCTCAGAAAGTTTGGGTGGGAGTTGACGCTACATCAGCGCTGCAAGTCTACTCCTC 65
 Qy 368 ThrGlyIleValIlePhePheAlaIleuValIlePheArgAsnGlnAlaSerGlyIleuValAla 387
 Db 66 ACAGAGATTGGGCTCTCTCTACTGATTTTGTGTGAGCCATTGACTCTCTATAGACC 125
 Qy 388 MetPheGlyThrThrProleuMetIleuValIleValIleGlnAsnIleu 407
 Db 126 AAGTTTGGAGTACACCTTTGCTTGGCGGAGCTCATGTTGGAGCATGACACATTTTC 185
 Qy 408 SerIlySerThirlyTyAlaIleuPheAspSerThirlysglMetAlaTyrllePro 426
 Db 186 AGTAAAGTGCAGAAATACAGTCTGTTGATCCTTGCAAAAGAGATGCATCATTCT 242
 RESULT 13
 ABX86909
 ID ABX86909 standard; cDNA; 280 BP.
 XX
 AC ABX86909;
 XX
 DT 24-APR-2003 (first entry)
 XX
 DE Corn ear-derived polynucleotide (cpd) #5369.
 XX
 KM Corn ear-derived polynucleotide; cpd; cDNA library; SATMON022;
 KM SATMON023; structural gene; functional gene; regulatory gene;
 KM corn ear-specific profile; gene transcription; gene expression;
 KM hybrid plant; desirable trait expression; plant breeding program;
 KM inheritance; desired characteristic; growth; development;
 KM disease resistance; environmental adaptability; quality; yield;
 KM multigene trait; plant; gene; ss.
 XX
 OS Zea mays.
 XX
 PN US6476212-B1.
 XX
 PD 05-NOV-2002.
 XX
 PF 14-MAY-1999; 99US-0313294.
 XX
 PR 26-MAY-1998; 98US-086722P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lalgudi RV; Ito LY, Sherman BK;
 DR WPI; 2003-208840/20.
 XX
 PT Novel purified corn-ear derived polynucleotide useful as hybridization
 PT probe for detecting polynucleotide in sample, and for identifying,
 PT evaluating, and altering desired characteristics associated with
 PT growth, development -
 XX
 PS Examples; SEQ ID No 5369; 390pp; English.
 XX
 CC The present invention relates to the isolation of corn ear-derived

CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries
CC SAMON022 and SAMON023. Some of the cpds uniquely identify structural,
CC functional, and regulatory genes of corn ear. The polynucleotides
CC sequences are useful for detecting cpds in a sample, for producing
CC a corn ear-specific profile of gene transcription, for detecting
CC altered gene expression in inbred or hybrid plants, and for screening
CC several molecules for specific binding to the polynucleotide. The cpds
CC are useful to identify isolate, or extend identical or related
CC corn ear nucleic acid sequences from DNA libraries, and in nucleic
CC acid amplification or hybridisation techniques to follow the
CC expression of desirable traits through plant breeding programs.
CC Preferably, the cpds are used to identify, evaluate, alter, or
CC follow the inheritance of desired characteristics associated with
CC growth and development, disease resistance, environmental adaptability,
CC quality, and yield of corn. The cpds are also useful as molecular
CC markers for studying inheritance and multigene traits in a plant
CC breeding program. The cpds are useful for producing purified corn-ear
CC polypeptides by recombinant techniques. They are also useful in
CC diagnostic assays to detect or confirm conditions or diseases
CC associated with abnormal levels of cdp expression. ABX81541-ABX89140
CC represent corn ear-derived polynucleotides (cpds) of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site at seqdata.uspto.gov/psipeditEntry.html.
XX

Sequence 280 BP; 65 A; 65 C; 74 G; 75 T; 1 other;

Alignment Scores:
Pred. No.: 3,396-17 Length: 280
Score: 242.00 Matches: 50
Percent Similarity: 77.33% Conservative: 8
Best Local Similarity: 66.67% Mismatches: 17
Query Match: 9,208 Indels: 0
DB: 25 Gaps: 0

US-09-869-433-2 (1-515) x ABX86909 (1-280)

QY 402 AaiaiegnaaiaileuSerlySerThryTyraiaLeupheAspSerThryyeglu 421
DB 3 GCAATGCGAATTTTCAGTAGAGTGCAGAAATACAGTCTTTGATCTTGCAGAG 62
QY 422 MetatyrilepLeuAspGlnGlnGlnValysGlyTyraiaaiaileaspval 441
DB 63 ATGGATATCATTTCTTGATGAGATATGAAGTGAAGGTAAGCGCATTTGATGTT 122
QY 442 ValaiaaiaargpneGlyLysSerGlyGlyalaLeuileGlnGlnGlyLeuValile 461
DB 123 GTGTGCAACCCCTTGGGAAATCTGAGGTGCTCTGATCCAGCATTCATCATCTGTCA 182
QY 462 CysGlySerileGlyAlaMetThryProTyraiaValileLeu 476
DB 183 TTCGTTCTCTCGCACTCGACACCTTACTTGGCGAAATATCTC 227

RESULT 14

ABL71938
ID ABL71938 standard; cDNA; 309 BP.

XX ABL71938;

XX 14-MAY-2002 (first entry)

DE Corn tassal-derived polynucleotide (cdps) SEQ ID NO:1312.

XX Corn; corn tassal-derived polynucleotide; cpds; hybrid breeding; CDPs;

KM inheritance; characteristic; growth; development; disease resistance;

KM environmental adaptability; quality; yield; molecular marker;

XX multigene trait; plant breeding; corn tassal; gene; ss.

OS Zea mays.

XX US2001051335-A1.

XX 13-DEC-2001.

XX 16-APR-1999; 99US-0294093.
PE 21-APR-1998; 98US-082567P.
FR (LALG/) LALGUDI R V.
XX (ITOL/) ITO L Y.
PA (SHER/) SHERMAN B K.
XX Lajgudi RV, Ito LY, Sherman BK;
XX WPI; 2002-163647/21.
DR
XX Novel purified corn tassal-derived polynucleotide useful for
PT determining altered gene expression, to recover regulatory elements and
PT to follow inheritance of desirable characteristics through hybrid
PT breeding programs -
XX

Claim 1; SEQ ID 1312; 201pp; English.

CC The present sequence describes a purified corn tassal-derived
CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
CC selected from those given in ABL70627 to ABL76833. The cpds sequences
CC encode corn tassal-derived polypeptides (CDPs). The cpds sequences (1)
CC can be used for determining altered gene expression, to recover
CC regulatory elements and to follow inheritance of desirable
CC characteristics through hybrid breeding programs. (1) are also useful
CC in the evaluation, and alteration of desired characteristics associated
CC with growth and development, disease resistance, environmental
CC adaptability, quality and yield, and as molecular markers for studying
CC inheritance of multigene traits in a plant breeding program. (1) can be
CC used to produce a tassal-specific profile of gene transcription, a
CC transcript image, to clone regulatory elements for use in transformation
CC vectors, to express a polypeptide, to identify, isolate or extend
CC identical or related corn tassal nucleic acid sequences from DNA
CC libraries, in nucleic acid hybridisation or amplification technologies,
CC as query sequences to determine homology of known sequences, as probe
CC for use in Southern or Northern hybridisation, and to identify the
CC presence of and/or to determine the degree of similarity between two
CC (or more) nucleic acid sequences.

XX Sequence 309 BP; 82 A; 52 C; 78 G; 97 T; 0 other;

Alignment Scores:
Pred. No.: 1,496-15 Length: 309
Score: 227.50 Matches: 46
Percent Similarity: 66.67% Conservative: 24
Best Local Similarity: 43.81% Mismatches: 32
Query Match: 8,658 Indels: 3
DB: 24 Gaps: 2

US-09-869-433-2 (1-515) x ABL71938 (1-309)

QY 183 ArgPheTyraiaLeuPheGlyileGlyAlaAsnileSerLeuMetAlaSerGlyArgAla 202
DB 1 AAATTCACCCGTTATTTGGCCTTGGGCTTAATATCTCTTCTTCTTGGCGACT 60
QY 203 IleValTrpAlaSerLysLeuArgAlaSerValSerGlnGlyValaAspProTrpGlyile 222
DB 61 GTGAAGTATTTCTCAATTTCCGACACACATTTGGCTCTGGAATGTAGTGGTGAAGTA 120
QY 223 SerLeuArgLeuMetAlaMetThryileValSerGlyLeuValleuMetAlaSerTyr 242
DB 121 TCTTTGAAGAGAAAGATGACATATGCTGCTTGGACCTGTGATCAATCTTCATATAT 180
QY 243 TrpTrpIleAsnLysAsnValleuThraspProArgPheTyraiaProGlnGlnMetGln 262
DB 181 TGGGAGTGAACAAGTTTGTTTGAATGATCTTCACTT-----CCAAAGGCTGATCGT 234
QY 263 LysGlyLysLysGlyAlaLysProLysMetAsnMetLysAspSerPheLeuTyraiaAsp 282
DB 235 AAG---AAGAAAAAGCAAAAGCCTAAACTTGATGAAAGAGATCGAAGGTTCTGCTC 291

QY 283 ArgSerProTyrIle 287
 DB 292 TCCTCGAGGTATG 306
 RESULT 15
 ID ABX84884 standard; cDNA, 265 BP.
 XX ABX84884;
 AC
 XX 24-APR-2003 (first entry)
 XX
 XX Corn ear-derived polynucleotide (cpd) #3344.
 XX
 XX Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022;
 KM SATMON023 structural gene; functional gene; regulatory gene;
 KM corn ear-specific profile; gene transcription; gene expression;
 KM hybrid plant; desirable trait expression; plant breeding program;
 KM inheritance; desired characteristic; growth; development;
 KM disease resistance; environmental adaptability; quality; yield;
 KM multigene trait; plant; gene; ss.
 XX
 XX Zea mays.
 OS
 XX US6476212-B1.
 PN
 XX 05-NOV-2002.
 PD
 XX 14-MAY-1999; 99US-0313294.
 PF
 XX 26-MAY-1998; 98US-086722P.
 PR
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 PI Laigudi RV, Ito LY, Sherman BK;
 PI WPI; 2003-208840/20.
 DR
 XX
 XX Novel purified corn-ear derived polynucleotide useful as hybridization
 PT probe for detecting polynucleotide in sample, and for identifying,
 PT evaluating, and altering desired characteristics associated with
 PT growth, development
 PT
 XX
 XX Examples; SEQ ID No 3344; 390pp; English.
 PS
 XX The present invention relates to the isolation of corn ear-derived
 CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries
 CC SATMON022 and SATMON023. Some of the cpds uniquely identify structural,
 CC functional, and regulatory genes of corn ear. The polynucleotides
 CC sequences are useful for detecting cpds in a sample, for producing
 CC a corn ear-specific profile of gene transcription, for detecting
 CC altered gene expression in inbred or hybrid plants, and for screening
 CC several molecules for specific binding to the polynucleotide. The cpds
 CC are useful to identify, isolate, or extend identical or related
 CC corn-ear nucleic acid sequences from DNA libraries, and in nucleic
 CC acid amplification or hybridisation techniques to follow the
 CC expression of desirable traits through plant breeding programs.
 CC Preferably, the cpds are used to identify, evaluate, alter, or
 CC follow the inheritance of desired characteristics associated with
 CC growth and development, disease resistance, environmental adaptability,
 CC quality, and yield of corn. The cpds are also useful as molecular
 CC markers for studying inheritance and multigene traits in a plant
 CC breeding program. The cpds are useful for producing purified corn-ear
 CC polypeptides by recombinant techniques. They are also useful in
 CC diagnostic assays to detect or confirm conditions or diseases
 CC associated with abnormal levels of cdp expression. ABX81541-ABX89140
 CC represent corn ear-derived polynucleotides (cpds) of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site at seqdata.uspto.gov/psipdIDEntry.html.
 XX
 SQ Sequence 265 BP; 73 A; 57 C; 60 G; 73 T; 2 other;

Alignment Scores:
 Pred. No.: 3,67e-09 Length: 265
 Score: 168.00 Matches: 34
 Percent Similarity: 76.92% Conservative: 6
 Best Local Similarity: 65.38% Mismatches: 12
 Query Match: 6.39% Indels: 0
 DB: 25 Gaps: 0
 US-09-869-433-2 (1-515) x ABX84884 (1-265)
 QY 421 GlnMetAlaTyrIleProLeuAspGlnGlnIleValIleGlyIleValAlaIleAsp 440
 DB 106 GAAATGGCATATATCCCTTTCGATGAAGATATGAAGTAAGAAAGGACGACCTTAT 165
 QY 441 ValValAlaAlaArgPheGlyIleSerGlyValAlaLeuIleGlnGlnIleVal 460
 DB 166 GTTGTTTGCAACCCACCTGGGCAAGTCGGCGCGCACCTTATCCAAACAGTTGATGATCCTG 225
 QY 461 IleCysGlySerIleGlyAlaMetThrProTyrIleu 472
 DB 226 ACGTTGGNCCCTGGCAAAATTCGACTTCATTCCTT 261

Search completed: November 25, 2003, 10:51:53
 Job time : 1796 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:13:32 ; Search time 22 Seconds

(Without alignments)
990,459 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 2630

Sequence: 1 MTETKEPKFKLRSFLMPFH.....AQSLAKQEVAGQSDAPASS 515

Scoring table:

BIOSDMM62

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/1/iaa/6C_COMB.pep:*
- 6: /cgn2_6/ptodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | Match | Length | ID | Description |
|------------|--------|-------|-------|--------|----------------------|--------------------|
| 1 | 2617 | 99.5 | 524 | 4 | US-09-198-452A-369 | Sequence 369, App |
| 2 | 1025.5 | 39.0 | 551 | 4 | US-09-198-452A-653 | Sequence 653, App |
| 3 | 157.5 | 6.0 | 907 | 4 | US-09-198-452A-306 | Sequence 306, App |
| 4 | 135 | 5.1 | 545 | 4 | US-09-252-991A-32219 | Sequence 32219, A |
| 5 | 129 | 4.9 | 488 | 4 | US-09-252-991A-17902 | Sequence 17902, A |
| 6 | 122.5 | 4.7 | 713 | 3 | US-09-335-409-11 | Sequence 11, App1 |
| 7 | 122.5 | 4.7 | 713 | 4 | US-09-568-102-11 | Sequence 11, App1 |
| 8 | 122.5 | 4.7 | 713 | 4 | US-09-567-969-11 | Sequence 11, App1 |
| 9 | 122.5 | 4.7 | 713 | 4 | US-09-568-480-11 | Sequence 11, App1 |
| 10 | 122.5 | 4.7 | 713 | 4 | US-09-568-472-11 | Sequence 11, App1 |
| 11 | 122.5 | 4.7 | 713 | 4 | US-09-567-899-11 | Sequence 11, App1 |
| 12 | 122.5 | 4.7 | 713 | 4 | US-09-252-991A-30725 | Sequence 30725, A |
| 13 | 120.5 | 4.6 | 534 | 4 | US-09-252-991A-32764 | Sequence 32764, A |
| 14 | 120.5 | 4.6 | 1162 | 4 | US-09-134-001C-5655 | Sequence 5655, App |
| 15 | 119.5 | 4.5 | 800 | 4 | US-09-328-352-8165 | Sequence 8165, App |
| 16 | 119 | 4.5 | 940 | 4 | US-09-107-532A-6563 | Sequence 6563, App |
| 17 | 115.5 | 4.4 | 546 | 4 | US-09-252-991A-17104 | Sequence 17104, A |
| 18 | 115.5 | 4.4 | 546 | 4 | US-09-328-352-6961 | Sequence 6961, App |
| 19 | 114 | 4.3 | 414 | 4 | US-09-134-001C-4637 | Sequence 4637, App |
| 20 | 113 | 4.3 | 578 | 4 | US-09-252-991A-18288 | Sequence 18288, A |
| 21 | 113 | 4.3 | 1095 | 3 | US-09-112-086-15 | Sequence 15, App1 |
| 22 | 113 | 4.3 | 419 | 4 | US-09-328-352-974 | Sequence 974, App |
| 23 | 112.5 | 4.3 | 360 | 1 | US-08-597-236-11 | Sequence 11, App1 |
| 24 | 111 | 4.2 | 360 | 1 | US-08-746-682A-11 | Sequence 11, App1 |
| 25 | 111 | 4.2 | 770 | 4 | US-09-252-991A-25267 | Sequence 25267, A |
| 26 | 111 | 4.2 | 770 | 4 | US-09-107-532A-6105 | Sequence 6105, App |
| 27 | 110 | 4.2 | 478 | 4 | US-09-107-532A-6105 | Sequence 6105, App |

| | | | | | | |
|-----|-------|-----|------|---|----------------------|--------------------|
| 28 | 109.5 | 4.2 | 540 | 4 | US-09-107-532A-5467 | Sequence 5467, App |
| 29 | 109 | 4.1 | 478 | 4 | US-09-328-352-7771 | Sequence 7771, App |
| 30 | 108.5 | 4.1 | 548 | 4 | US-09-328-352-6605 | Sequence 6605, App |
| 31 | 108.5 | 4.1 | 760 | 4 | US-09-555-313B-2 | Sequence 2, App1 |
| 32 | 108.5 | 4.1 | 816 | 4 | US-09-252-991A-24105 | Sequence 24105, A |
| 33 | 108 | 4.1 | 463 | 4 | US-09-252-991A-29935 | Sequence 29935, A |
| 34 | 107.5 | 4.1 | 499 | 4 | US-09-328-352-5089 | Sequence 5089, App |
| 35 | 107 | 4.1 | 417 | 4 | US-09-328-352-4168 | Sequence 4168, App |
| 36 | 106 | 4.0 | 512 | 4 | US-09-107-532A-5262 | Sequence 5262, App |
| 37 | 105.5 | 4.0 | 360 | 4 | US-09-555-313B-4 | Sequence 4, App1 |
| 38 | 105.5 | 4.0 | 380 | 4 | US-09-555-313B-24 | Sequence 24, App1 |
| 39 | 105.5 | 4.0 | 388 | 1 | US-08-446-822-8 | Sequence 8, App1 |
| 40 | 105.5 | 4.0 | 388 | 1 | US-09-328-352-14-8 | Sequence 8, App1 |
| 41 | 105.5 | 4.0 | 388 | 5 | PCT-US93-12586-8 | Sequence 8, App1 |
| 42 | 105.5 | 4.0 | 366 | 4 | US-09-134-001C-3355 | Sequence 3355, App |
| 43 | 105 | 4.0 | 569 | 2 | US-08-750-723A-2 | Sequence 2, App1 |
| 44 | 105 | 4.0 | 569 | 3 | US-09-191-875-2 | Sequence 2, App1 |
| 45 | 104.5 | 4.0 | 483 | 4 | US-09-107-532A-4123 | Sequence 4123, App |
| 46 | 104 | 4.0 | 1160 | 4 | US-09-328-352-6826 | Sequence 6826, App |
| 47 | 103.5 | 3.9 | 443 | 4 | US-09-134-001C-5557 | Sequence 5557, App |
| 48 | 103.5 | 3.9 | 443 | 4 | US-09-328-352-7567 | Sequence 7567, App |
| 49 | 103 | 3.9 | 503 | 4 | US-09-252-991A-32663 | Sequence 32663, A |
| 50 | 103 | 3.9 | 776 | 4 | US-09-165-396-3 | Sequence 3, App1 |
| 51 | 102.5 | 3.9 | 439 | 4 | US-09-328-352-7206 | Sequence 7206, App |
| 52 | 102.5 | 3.9 | 901 | 4 | US-09-134-001C-5389 | Sequence 5389, App |
| 53 | 102.5 | 3.9 | 477 | 4 | US-09-252-991A-19515 | Sequence 19515, A |
| 54 | 102 | 3.9 | 477 | 4 | US-09-134-001C-3467 | Sequence 3467, App |
| 55 | 101 | 3.8 | 385 | 4 | US-09-328-352-5641 | Sequence 5641, App |
| 56 | 101 | 3.8 | 512 | 4 | US-09-328-352-6192 | Sequence 6192, App |
| 57 | 101 | 3.8 | 521 | 4 | US-09-107-532A-3961 | Sequence 3961, App |
| 58 | 100.5 | 3.8 | 234 | 4 | US-09-107-532A-3787 | Sequence 3787, App |
| 59 | 100 | 3.8 | 335 | 4 | US-09-328-352-4577 | Sequence 4577, App |
| 60 | 100 | 3.8 | 335 | 4 | US-09-252-991A-25524 | Sequence 25524, A |
| 61 | 100 | 3.8 | 413 | 4 | US-09-134-001C-5503 | Sequence 5503, App |
| 62 | 100 | 3.8 | 527 | 4 | US-09-252-991A-17545 | Sequence 17545, A |
| 63 | 99.5 | 3.8 | 2005 | 3 | US-08-836-325-7 | Sequence 7, App1 |
| 64 | 99 | 3.8 | 425 | 4 | US-09-328-352-7964 | Sequence 7964, App |
| 65 | 99 | 3.8 | 453 | 4 | US-09-252-991A-17394 | Sequence 17394, A |
| 66 | 99 | 3.8 | 600 | 4 | US-09-252-991A-22396 | Sequence 22396, A |
| 67 | 98.5 | 3.7 | 503 | 4 | US-09-134-001C-3948 | Sequence 3948, App |
| 68 | 98.5 | 3.7 | 598 | 2 | US-08-853-659A-53 | Sequence 53, App1 |
| 69 | 98.5 | 3.7 | 831 | 2 | US-08-677-734A-11 | Sequence 11, App1 |
| 70 | 98.5 | 3.7 | 831 | 4 | US-09-097-053-11 | Sequence 11, App1 |
| 71 | 98 | 3.7 | 403 | 4 | US-09-328-352-7791 | Sequence 7791, App |
| 72 | 98 | 3.7 | 422 | 4 | US-09-134-001C-4048 | Sequence 4048, App |
| 73 | 98 | 3.7 | 473 | 4 | US-09-252-991A-28845 | Sequence 28845, A |
| 74 | 97 | 3.7 | 443 | 4 | US-09-328-352-7069 | Sequence 7069, App |
| 75 | 97 | 3.7 | 445 | 4 | US-09-252-991A-33142 | Sequence 33142, A |
| 76 | 97 | 3.7 | 502 | 4 | US-09-328-352-6968 | Sequence 6968, App |
| 77 | 96.5 | 3.7 | 274 | 4 | US-09-489-847-278 | Sequence 278, App |
| 78 | 96.5 | 3.7 | 449 | 4 | US-09-328-352-7512 | Sequence 7512, App |
| 79 | 96.5 | 3.7 | 482 | 4 | US-09-134-001C-4309 | Sequence 4309, App |
| 80 | 96.5 | 3.7 | 483 | 4 | US-09-252-991A-21015 | Sequence 21015, A |
| 81 | 96 | 3.7 | 418 | 4 | US-09-252-991A-30786 | Sequence 30786, A |
| 82 | 96 | 3.7 | 462 | 4 | US-09-328-352-6888 | Sequence 6888, App |
| 83 | 96 | 3.7 | 476 | 4 | US-09-328-352-5027 | Sequence 5027, App |
| 84 | 95.5 | 3.6 | 407 | 4 | US-09-252-991A-21511 | Sequence 21511, A |
| 85 | 95.5 | 3.6 | 559 | 4 | US-09-328-352-5924 | Sequence 5924, App |
| 86 | 95.5 | 3.6 | 585 | 4 | US-09-328-352-6133 | Sequence 6133, App |
| 87 | 95.5 | 3.6 | 623 | 4 | US-09-252-991A-19867 | Sequence 19867, A |
| 88 | 95.5 | 3.6 | 793 | 4 | US-09-107-532A-5141 | Sequence 5141, App |
| 89 | 95 | 3.6 | 364 | 4 | US-09-134-001C-4201 | Sequence 4201, App |
| 90 | 95 | 3.6 | 420 | 4 | US-09-107-532A-4006 | Sequence 4006, App |
| 91 | 95 | 3.6 | 450 | 4 | US-09-252-991A-25019 | Sequence 25019, A |
| 92 | 95 | 3.6 | 451 | 4 | US-09-328-352-5244 | Sequence 5244, App |
| 93 | 95 | 3.6 | 466 | 4 | US-09-328-352-4381 | Sequence 4381, App |
| 94 | 95 | 3.6 | 502 | 4 | US-09-252-991A-22327 | Sequence 22327, A |
| 95 | 95 | 3.6 | 813 | 3 | US-08-836-325-8 | Sequence 8, App1 |
| 96 | 95 | 3.6 | 865 | 4 | US-09-252-991A-19339 | Sequence 19339, A |
| 97 | 94.5 | 3.6 | 242 | 4 | US-09-489-847-134 | Sequence 134, App |
| 98 | 94.5 | 3.6 | 321 | 4 | US-09-688-019-2 | Sequence 2, App1 |
| 99 | 94.5 | 3.6 | 395 | 4 | US-09-328-352-5819 | Sequence 5819, App |
| 100 | 94.5 | 3.6 | 400 | 4 | US-09-198-452A-3302 | Sequence 3302, App |

ALIGNMENTS

RESULT 1
US-09-198-452A-369

; Sequence 369, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifflais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 369
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-369

Query Match 99.5%; Score 2617; DB 4; Length 524;
Best Local Similarity 99.6%; Pred. No. 6.4e-274;
Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY      1  MTKTEKPKGLRSLFMPHITHELKVLPMFLMFCITFNTYTVLRDTKDTLIYVAPSGA 60
DB      10  MTKTEKPKGLRSLFMPHITHELKVLPMFLMFCITFNTYTVLRDTKDTLIYVAPSGA 69
QY      61  EAIPIKFWLVPCALIFMLIYAKLSNLSKQALFYANGTPELIFPFIYIPLRDVL 120
DB      70  EAIPIKFWLVPCALIFMLIYAKLSNLSKQALFYANGTPELIFPFIYIPLRDVL 129
QY     121  HPTFADRLQALIPPGILGLVALILRNMTFAFYVLAELMGSVMLSLMFWGFANEITKHE 180
DB     130  HPTFADRLQALIPPGILGLVALILRNMTFAFYVLAELMGSVMLSLMFWGFANEITKHE 189
QY     181  AKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLMAAMTIVSGVLMA 240
DB     190  AKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLMAAMTIVSGVLMA 249
QY     241  SYWINKNVLTDPFRYPNEEMQKKGAKPKKMKMDSTLYIDRSYIILLTLVIAYGIC 300
DB     250  SYWINKNVLTDPFRYPNEEMQKKGAKPKKMKMDSTLYIDRSYIILLTLVIAYGIC 309
QY     301  INLEVTAKSOLKQYPMNDYSEFMGNFSFWTGVSVLIMLVGAGNVIKFGMLTGALV 360
DB     310  INLEVTAKSOLKQYPMNDYSEFMGNFSFWTGVSVLIMLVGAGNVIKFGMLTGALV 369
QY     361  TPVVVLTGIVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKYALPDSK 420
DB     370  TPVVVLTGIVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKYALPDSK 429
QY     421  EMATIPLDQEKVKGAIDVVAARFGSGGALIOQGLVTCGSIGANTPYLAIVLFI 480
DB     430  EMATIPLDQEKVKGAIDVVAARFGSGGALIOQGLVTCGSIGANTPYLAIVLFI 489
QY     481  AIWLVSATKLNKFLAOSALKOEVAQEDSAPASS 515
DB     490  AIWLVSATKLNKFLAOSALKOEVAQEDSAPASS 524

```

RESULT 2

US-09-198-452A-653
; Sequence 653, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifflais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention

TITLE OF INVENTION: and treatment of infection

FILE REFERENCE: 9710-003-999

CURRENT APPLICATION NUMBER: US/09/198,452A

CURRENT FILING DATE: 1998-11-24

NUMBER OF SEQ ID NOS: 6849

SEQ ID NO 653

LENGTH: 551

TYPE: PRT

ORGANISM: Chlamydia pneumoniae

US-09-198-452A-653

Query Match 39.0%; Score 1025.5; DB 4; Length 551;
Best Local Similarity 39.4%; Pred. No. 6e-102;
Matches 214; Conservative 107; Mismatches 185; Indels 37; Gaps 9;

```

QY      1  MTKTEKPKGLRSLFMPHITHELKVLPMFLMFCITFNTYTVLRDTKDTLIYVAPSGA 60
DB      12  MOSSEVPFSLRAVLCPIYKSEFSKFPVPLFLAFPGFVNCILKNNKDTLIVIGSDAGA 71
QY     61  EAIPIKFWLVPCALIFMLIYAKLSNLSKQALFYANGTPELIFPFIYIPLRDVL 120
DB     72  EAIPIKFWLVPCALIFMLIYAKLSNLSKQALFYANGTPELIFPFIYIPLRDVL 131
QY     121  HPTFADRLQALIPPGILGLVALILRNMTFAFYVLAELMGSVMLSLMFWGFANEITKHE 180
DB     132  HUNSLADKLQELPQGLRGFIWVRYSYISYMSLSMSVSLMFLWGLANGITITTE 191
QY     181  AKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLMAAMTIVSGVL 238
DB     192  AGRFYALINTGLNLSICAGEISYWMGKQTFVAYSFACDSWHSVWMLNLTMLIT-CGSLIM 250
QY     239  MASWYINKNV--LTDRFNPNEEMQKKGKGA-----KPKMKKDSFLYIDRSPIY 287
DB     251  I---WYKRHHHTIDTISIPSRRYLAEGATATANKKKKPKAKRNLFHLIQRYL 306
QY     288  LLTLVIAYGICINLEVTAKSOLKQYPMNDYSEFMGNFSFWTGVSVLIMLVGAGN 347
DB     307  IGLAIVLSYLVNVLHLEVMWQKQVQSQISHSHEFNMYMSITLLIGVSVLAIVLTLGQ 366
QY     348  VIRKFGMLTALVTPVVVLTGIVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONI 406
DB     367  CIRKMGWTVGALVTPMLVLSGLLFFGTIFPAKADISIFGIVLGMTPLAALAMTGKQNV 426
QY     407  LSKSTKALPDSKTEMAYIPLDQEKVKGAIDVVAARFGSGGALIOQGLVTCGSIG 466
DB     427  LSKSTKTFPDQTEMAYIPLDQEKVKGAIDVVAARFGSGGALIOQGLVTCGSIG 486
QY     467  AMTPYLAIVLFIATWLVSATKLNKFLAOSA-----LK-----EQEVAQE 508
DB     487  ASLVNIALVLLIIMVWIAVAVAYIGKEYGRADAQVATLQPKPSSSIYREAQESVQE 546
QY     509  DSA 511
DB     547  EMA 549

```

RESULT 3

US-09-198-452A-306
; Sequence 306, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifflais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 306
; LENGTH: 907
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae

US-09-198-452A-306

| | | | | |
|-----------------------|------------------|--------------------|-------------|-------------|
| Query Match | 6.0%; | Score 157.5; | DB 4; | Length 907; |
| Best Local Similarity | 20.4%; | Pred. No. 7.5e-08; | | |
| Matches 94; | Conservative 95; | Mismatches 163; | Indels 109; | Gaps 26 |

[illegible]

```

RESULT 4
US-09-252-991A-32219
; Sequence 32219 Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32219
; LENGTH: 545
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-32219

```

Query Match 5.1%; Score 135; DB 4; Length 545;
Best Local Similarity 21.1%; Pred. No. 9, 5e-06;
Matches 110; Conservative 58; Mismatches 163; Indels 190; Gaps 26

| | | | |
|----|-----|--|-----|
| QY | 32 | LMPECFINPVAVDRTKDTLVG-----AAGSGEAAPF--IKRWLVPCALIMLYA | 83 |
| | | | |
| DB | 143 | LVFPAITLNTI-----DRAAGVWQPLIAEMSTANDYANINPFQVGNIGETLIGR | 196 |
| | | | |
| QY | 84 | KLSNLSKQALFYAVGTFLLFPALFPVIVYPLRDVHPTEPADRLQAILPGLIGLVAI | 143 |
| | | | |
| DB | 197 | FIDRVGVRRAEFLV-----LWLSLTGAGLATS | 226 |

```

QY 144 LRMTFAAFY-LAEI--WGSVMISLMFWGPNELTKIHEARFPA--LEFGANISLLA 198
Db 227 AAGFWCRFLLGLTEANYPACVKTRLMFPGE-----RAVATGFMAGTNGAAY 278
QY 199 SGRAI-----WASKLRASVEGVDPMGISRLMAMTVISGLVIMASYMINKVLTDP 253
Db 279 TPALLPLLSWVG-----WQAFLOWGSL--GLWMLV-FMWN- 313
QY 254 RPYNDEKQKKG-----AKPKNMKDSFLYLDPSYILLTLTLVIA----- 296
Db 314 -YNPBEHPRVKQSELEYIQOOBEPATRVPSQILRRRGTAFAFASITPAVFEWFLY 372
QY 297 -----YGCINLIEVTWKSQKLOKYPMNNDVSEFMGNFSFMTGVSVLIML----- 342
Db 373 WLPPFLNOQYGGISLVTOM-----GIPLLIMLTADPG 405
QY 343 FVGGNVIRKPFMLTGCALVTPLYMVLITGLVFPAL-----YIFRNQASGL-VAMEGTPPLMLA 397
Db 406 SVGGIIS--SMLIGRGMFAIRARLISMLTFACTIVGVVFAANASGLWTAV-----LATA 458
QY 398 VVGAIQ-----NLSKSTKYALFDSTKEMAYIPLDOEKVYKAAIDVVAARFG-----XS 449
Db 459 LAVGAHQMTANIWS-----LVMDYTPKIMSTVF-----FGGMCAAL 497
QY 450 GGALIQGLVICSIGAMTPYLANILFLTAIMVSAFTKL 490
Db 498 GGMFPTQ---IVGGVLTATNNNVAALFTMIPIMMYTIALTWL 535

```

```

RESULT 5
US-09-252-991A-17902
: Sequence 17902, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: PRIOR FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 17902
: LENGTH: 488
: TYPE: PRF
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17902

```

```

Db 187 GGYAASATGLVGLGOMILVAHQGLPELMLMEVAFCEPALCLVPLALTHKIHAPALPAPLE 246
Qy 322 YSEPMGNF--SFWTGVUSVLM-----LFGVGNVRK 351
Db 247 PRFPIRVPOSILTYLVSLVSGLVGSFYGLAPYANQGLPNEQVGLYMGACIFAGLLVQMP 306
Qy 352 FGMVLTGA-----LVTVPVVLTLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONI 406
Db 307 LGWISDRDRAMLRACAILL---CLFALPL-----ALLQOMPLALLALGIAASM 354
Qy 407 LSKSTKVALPDKTEMAYIPLDQEQVKGAIDV--VAARFGK-SGGALIOGGLVTC 462
Db 355 L----QFTLYPLAFAFSDHETERVSLTMLVTFGVGACIGPLAAGALMR----- 403
Qy 463 GSIGAMPYLAIVLLFIATILVSAITKANKF----- 494
Db 404 -LFGANMLYAFVASCALILVWRHBEKVSGLHRVDAPLHHVPTPDNNTSSPLVAALDPR 462
Qy 495 LAQSALKEQVAVQEDSAP 512
Db 463 VDEQAVQEQVYDGEPPDAP 480

RESULT 6
US-09-335-409-11
; Sequence 11, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PR1
; ORGANISM: Sorangium cellulosum
US-09-335-409-11

Query Match 4.7%; Score 122.5; DB 3; Length 713;
Best Local Similarity 21.2%; Pred. No. 0.00032;
Matches 105; Conservative 67; Mismatches 165; Indels 159; Gaps 21;

Qy 76 IIFMLIYAKLSNLSKQ-----ALFYAVGTPLIFPALFPYIYPLRDVLPHTPEAD 127
Db 16 LALLIVTARASGELARLRQPEVLGELFGV-----VLGSPVAGLAPGFRALFOE 67
Qy 128 RLQAILPPGLGLGVALILRNWTFPAFYVLAELMGSVMLSMFWGFANBITKIHEAKRFYAL 187
Db 68 PAVGVVLSGI-----SWIGALLLLMAGIEVDVGIIRKARPPAL 107
Qy 68 PAVGVVLSGI-----SWIGALLLLMAGIEVDVGIIRKARPPAL 107
Db 188 FGIGANISLASGRAIYVASKLRASVSEGVDPWGISIRLIMAMTIVSGLV-----LM 239
Qy 108 SALGAIAPPLAAGA-----FSALVLDRLPLPSGLFGLIVSVTAIVSVIATVILIERESMR 161
Db 240 ASYMWIN--KNVLDPRFYNPEEMQKKGAKPKANMDSFLYIDRSFYIILLTLVAY 297
Qy 162 RSYAQVTLAAGVSEVAMVAVMNTSSSYGASPALAVARSALL--ASGFLIFMLV-- 215
Db 298 GICINLIEVTW-----KSQKLQYPMNDYSEFMGNFSFWTGVUSVLMLEFVGANVI 349
Qy 216 GRRLTHLAMRVADATRVSKQVSL-----VLVLTFLAALT 252
Db 350 RKFGW--LTGALVTPVNV-----LITGI-----VFPALVIFRNQASGL--V 386
Qy 253 QRLGHLPLGAFALGVLLNSAPRTNRPLLDGVQTLVAGLPAPVFFVLAKGVVDSQLRTP 312

```

```

Qy 367 AMEGTPEMLA-----VVGAIQNIILSKSTKVALPDKTEMAYIPLDQEQVKGAID 440
Db 313 AAMGTVALLLATATRAAVVPALGARLG-----GLRSEALVAVGLNM-----KGTD 361
Qy 441 VVAARFGKSGGALIOGGLVYICSGIGAMPYLAIVLLFIATL-----IMVSATKANKLFL 495
Db 362 LIVAIVG-----VELGILL-----SNEAYTMAYAVALTVAASPALLIMLE----- 401
Qy 496 AQSALKEQVAVQEDSA 511
Db 402 -----KRAPPOEBSA 412

RESULT 7
US-09-568-102-11
; Sequence 11, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PR1
; ORGANISM: Sorangium cellulosum
US-09-568-102-11

Query Match 4.7%; Score 122.5; DB 4; Length 713;
Best Local Similarity 21.2%; Pred. No. 0.00032;
Matches 105; Conservative 67; Mismatches 165; Indels 159; Gaps 21;

Qy 76 IIFMLIYAKLSNLSKQ-----ALFYAVGTPLIFPALFPYIYPLRDVLPHTPEAD 127
Db 16 LALLIVTARASGELARLRQPEVLGELFGV-----VLGSPVAGLAPGFRALFOE 67
Qy 128 RLQAILPPGLGLGVALILRNWTFPAFYVLAELMGSVMLSMFWGFANBITKIHEAKRFYAL 187
Db 68 PAVGVVLSGI-----SWIGALLLLMAGIEVDVGIIRKARPPAL 107
Qy 68 PAVGVVLSGI-----SWIGALLLLMAGIEVDVGIIRKARPPAL 107
Db 188 FGIGANISLASGRAIYVASKLRASVSEGVDPWGISIRLIMAMTIVSGLV-----LM 239
Qy 108 SALGAIAPPLAAGA-----FSALVLDRLPLPSGLFGLIVSVTAIVSVIATVILIERESMR 161
Db 240 ASYMWIN--KNVLDPRFYNPEEMQKKGAKPKANMDSFLYIDRSFYIILLTLVAY 297
Qy 162 RSYAQVTLAAGVSEVAMVAVMNTSSSYGASPALAVARSALL--ASGFLIFMLV-- 215
Db 298 GICINLIEVTW-----KSQKLQYPMNDYSEFMGNFSFWTGVUSVLMLEFVGANVI 349
Qy 216 GRRLTHLAMRVADATRVSKQVSL-----VLVLTFLAALT 252
Db 350 RKFGW--LTGALVTPVNV-----LITGI-----VFPALVIFRNQASGL--V 386
Qy 387 AMEGTPEMLA-----VVGAIQNIILSKSTKVALPDKTEMAYIPLDQEQVKGAID 440
Db 313 AAMGTVALLLATATRAAVVPALGARLG-----GLRSEALVAVGLNM-----KGTD 361
Qy 441 VVAARFGKSGGALIOGGLVYICSGIGAMPYLAIVLLFIATL-----IMVSATKANKLFL 495

```

Db 362 LIVAIVG-----VELGLL-----SNEAYTMAVVALVTVASPALLIMLE----- 401
QY 496 AQSALKEQEVNOEDSA 511
Db 402 -----KRAPPTOESBA 412

RESULT 8

US-09-567-969-11
; Sequence 11, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Jettan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyt, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIOR FILING DATE: 2000-05-10
; PRIOR FILING DATE: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-969-11

Query Match 4.7%; Score 122.5; DB 4; Length 713;

Best Local Similarity 21.2%; Pred. No. 0.00032;
Matches 105; Conservative 67; Mismatches 165; Indels 159; Gaps 21;

QY 76 IIFMLIVAKLSNLSKQ-----ALFYAVGTPPLIFPFFLPYIVPLRDVLPHTERAD 127
Db 16 LALIVTARASGELARLRQPEVGLFCGV-----VLGSPVGLAGPFRHALFOE 67
QY 128 RLQAILPPLGLGLVAILRNMTFAAFVYLAELMGSVMLSMFWGPNANETTKHEAKRFYAL 187
Db 68 PAVGVLSGI-----SWIGALLLLMAGIEVDVGIIRKEARPGAL 107
QY 188 FGIGANISLASGRAIIVASKLIRASVSGVDPWGISLRLMAMTIVSGLV-----LM 239
Db 108 SALGAIAPPLAAGA-----FSALVLDRLPLPSGLFGLIVSVTAIVSIYAKVIERESMR 161
QY 240 ASYWMIN--KNVLTDPREFYNEEMQKKGAKPKNMKDSFLYDRSPYILLTLVLVAY 297
Db 162 RSYAQVTLAAGVSEVAAWLVAMTSSSYGASPALAVARSALL--ASGFLFPMVLV----- 215
QY 298 GICINLIEVTW-----KSQKLQYNNMDYSEFMGNFSFWTGVSVLIMLFVGGNVI 349
Db 216 GRRLTHLAMRWVADATRVSKQVSL-----VLVLTFLAALLT 252
QY 350 RKFGM--LTGALVTPVWV-----LITGI-----VFPALVIFRNQASGL--V 386
Db 253 QRLGHPPLGAFALGVLLNSAPRTNRPLLDGVQTLVAGLPAPVFPVLGMRDVDSQLRTP 312
QY 387 AMFGTTPMLA-----VVVGAIONISLSTKYALFSTKEMAYPLDQEQVKGAID 440
Db 313 AAKGTVALLLATATRAKVPALGARLG-----GLRSEBALVAVGLNM-----KGTD 361
QY 441 VVAARFGKSGGALLQOGLVIVCGSIGAMTPYLAIVLLFTIA-----IMVSATKLNKFL 495
Db 362 LIVAIVG-----VELGIL-----SNEAYTMAVVALVTVASPALLIMLE----- 401
QY 496 AQSALKEQEVNOEDSA 511
Db 402 -----KRAPPTOESBA 412

RESULT 9

US-09-568-480-11
; Sequence 11, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Jettan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyt, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIOR FILING DATE: 2000-05-10
; PRIOR FILING DATE: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-11

Query Match 4.7%; Score 122.5; DB 4; Length 713;

Best Local Similarity 21.2%; Pred. No. 0.00032;
Matches 105; Conservative 67; Mismatches 165; Indels 159; Gaps 21;

QY 76 IIFMLIVAKLSNLSKQ-----ALFYAVGTPPLIFPFFLPYIVPLRDVLPHTERAD 127
Db 16 LALIVTARASGELARLRQPEVGLFCGV-----VLGSPVGLAGPFRHALFOE 67
QY 128 RLQAILPPLGLGLVAILRNMTFAAFVYLAELMGSVMLSMFWGPNANETTKHEAKRFYAL 187
Db 68 PAVGVLSGI-----SWIGALLLLMAGIEVDVGIIRKEARPGAL 107
QY 188 FGIGANISLASGRAIIVASKLIRASVSGVDPWGISLRLMAMTIVSGLV-----LM 239
Db 108 SALGAIAPPLAAGA-----FSALVLDRLPLPSGLFGLIVSVTAIVSIYAKVIERESMR 161
QY 240 ASYWMIN--KNVLTDPREFYNEEMQKKGAKPKNMKDSFLYDRSPYILLTLVLVAY 297
Db 162 RSYAQVTLAAGVSEVAAWLVAMTSSSYGASPALAVARSALL--ASGFLFPMVLV----- 215
QY 298 GICINLIEVTW-----KSQKLQYNNMDYSEFMGNFSFWTGVSVLIMLFVGGNVI 349
Db 216 GRRLTHLAMRWVADATRVSKQVSL-----VLVLTFLAALLT 252
QY 350 RKFGM--LTGALVTPVWV-----LITGI-----VFPALVIFRNQASGL--V 386
Db 253 QRLGHPPLGAFALGVLLNSAPRTNRPLLDGVQTLVAGLPAPVFPVLGMRDVDSQLRTP 312
QY 387 AMFGTTPMLA-----VVVGAIONISLSTKYALFSTKEMAYPLDQEQVKGAID 440
Db 313 AAKGTVALLLATATRAKVPALGARLG-----GLRSEBALVAVGLNM-----KGTD 361
QY 441 VVAARFGKSGGALLQOGLVIVCGSIGAMTPYLAIVLLFTIA-----IMVSATKLNKFL 495
Db 362 LIVAIVG-----VELGIL-----SNEAYTMAVVALVTVASPALLIMLE----- 401
QY 496 AQSALKEQEVNOEDSA 511
Db 402 -----KRAPPTOESBA 412

RESULT 10

US-09-568-486-11
; Sequence 11, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James

```

; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goetlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PR
; ORGANISM: Sorangium cellulosum
; US-09-568-486-11

Query Match      4.7%; Score 122.5; DB 4; Length 713;
Best Local Similarity 21.2%; Pred. No. 0.00032;
Matches 105; Conservative 67; Mismatches 165; Indels 159; Gaps 21;

QY 76 IIFMLIYAKLSNLSKQ-----ALFYAVGTPPLIFPAPFYIYPLRDVLPHTPEAD 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 16 LALILVTARASGELARLRQPEVLGELFGV-----VLGPSVVGALAPGFHRLPQE 67

QY 128 RLQALILPGLLGIVAILRNWTFAPFYVLAELMGVSLMFMGFANEITKHEAKRFYAL 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 68 PAVGVVLSGI-----SWIGALLLLLMAGIEVDVGIIRKEARPGAL 107

QY 188 FGIGANISILASGRAIWMASKLRASVSEGVDPWGISRLMAMTIVSGLV-----LM 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 108 SALGAIAPPLAAGAA-----FSALVLDRLPLPSGLFGIVLSTAVASVIKILIERESMR 161

QY 240 ASYWMIN--KNVLTDPREFYNPEBEMOKGKAKPKNMKDSFLYIDRSFYILLTLVLVAY 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 162 RSYAQVTLAAGVSEVAVWLVAMTSSSYGASPALAVARSALL--ASGFLLFMVLV--- 215

QY 298 GICINLIEVTW-----KSQKLQYPMNDYSEMGKFSFWTGVSVLIMLPFGGVNI 349
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 216 GRRLHLMRWVADATRVSKGVSL-----VLVTLFLAALT 252

QY 350 RKFGW--LTGALVTPVMV-----LTTGI-----VFPALVIFRQASGL--V 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 253 QRLGHPILGAFALGVILNSAPRTNRPLLDGVQTLVAGLPAPVFFVLAKMRVDVSQLRTP 312

QY 387 AMFGTTPMLA-----VVGAIONILSKSTKYALFDSKEMAYIPLDQOKYKGAID 440
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 313 AAMGTVALLLATATATAKVVPALGARLG-----GLRSEALVAVGLNM-----KGGTD 361

QY 441 VVAARFGSGGALLIOGGLLVICSGIGAMTPYLAIVLLFIIA-----IMVSAATKLNKFL 495
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 362 LIVAIVG-----VELGLL-----SNEAYTMVAVVAVLVITJASPALLIWLE----- 401

QY 496 AQSALKEQEVAAQEDSA 511
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 402 -----KRAPPTQESBA 412

RESULT 11
US-09-568-472-11
; Sequence 11, Application US/09568472
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goetlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PR
; ORGANISM: Sorangium cellulosum
; US-09-568-472-11
```

```

; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PR
; ORGANISM: Sorangium cellulosum
; US-09-568-472-11

Query Match      4.7%; Score 122.5; DB 4; Length 713;
Best Local Similarity 21.2%; Pred. No. 0.00032;
Matches 105; Conservative 67; Mismatches 165; Indels 159; Gaps 21;

QY 76 IIFMLIYAKLSNLSKQ-----ALFYAVGTPPLIFPAPFYIYPLRDVLPHTPEAD 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 16 LALILVTARASGELARLRQPEVLGELFGV-----VLGPSVVGALAPGFHRLPQE 67

QY 128 RLQALILPGLLGIVAILRNWTFAPFYVLAELMGVSLMFMGFANEITKHEAKRFYAL 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 68 PAVGVVLSGI-----SWIGALLLLLMAGIEVDVGIIRKEARPGAL 107

QY 188 FGIGANISILASGRAIWMASKLRASVSEGVDPWGISRLMAMTIVSGLV-----LM 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 108 SALGAIAPPLAAGAA-----FSALVLDRLPLPSGLFGIVLSTAVASVIKILIERESMR 161

QY 240 ASYWMIN--KNVLTDPREFYNPEBEMOKGKAKPKNMKDSFLYIDRSFYILLTLVLVAY 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 162 RSYAQVTLAAGVSEVAVWLVAMTSSSYGASPALAVARSALL--ASGFLLFMVLV--- 215

QY 298 GICINLIEVTW-----KSQKLQYPMNDYSEMGKFSFWTGVSVLIMLPFGGVNI 349
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 216 GRRLHLMRWVADATRVSKGVSL-----VLVTLFLAALT 252

QY 350 RKFGW--LTGALVTPVMV-----LTTGI-----VFPALVIFRQASGL--V 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 253 QRLGHPILGAFALGVILNSAPRTNRPLLDGVQTLVAGLPAPVFFVLAKMRVDVSQLRTP 312

QY 387 AMFGTTPMLA-----VVGAIONILSKSTKYALFDSKEMAYIPLDQOKYKGAID 440
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 313 AAMGTVALLLATATATAKVVPALGARLG-----GLRSEALVAVGLNM-----KGGTD 361

QY 441 VVAARFGSGGALLIOGGLLVICSGIGAMTPYLAIVLLFIIA-----IMVSAATKLNKFL 495
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 362 LIVAIVG-----VELGLL-----SNEAYTMVAVVAVLVITJASPALLIWLE----- 401

QY 496 AQSALKEQEVAAQEDSA 511
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 402 -----KRAPPTQESBA 412

RESULT 12
US-09-567-899-11
; Sequence 11, Application US/09567899
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goetlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
```

TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-869-11

Query Match 4.7%; Score 122.5; DB 4; Length 713;
Best Local Similarity 21.2%; Pred. No. 0.0032;
Matches 105; Conservative 67; Mismatches 165; Indels 159; Gaps 21;

Qy 76 IIFMLIVAKLSNLSKQ-----ALFYAVGTPPLIFALFPYIYPLRVNHPTEPAD 127
Db 16 LALLVTRASGELARLRQPEVLGELRGV-----VLGPSVGLAAGFPHALPQE 67
Qy 128 RLQALPFGGLGLVAIILNMTFAAFVYLAELMGSVLSLMFGFANEITKIHAKRFPAL 187
Db 68 PAVGVLSGI-----SWIGALLILMAGLEVVDGIIIRKARFAL 107
Qy 188 FGIGANISLASGRAIYVASKLRASVSEGVDPWGISLRLLAMNTVSGLV-----LM 239
Db 108 SALGAIAPPLAAGAA-----FSALVLDRLPLPSGLFLGIVSVTVASVIAKVLIERESMR 161
Qy 240 ASYMMIN--KNVLDPRFYNPEEMOKKAKPKPMKMSFLYIDRSFYILLTLVIAY 297
Db 162 RSYQVTLAAGVSEVAAVLMVAMTSSSYGASPALVARSALL--ASGFLFMVLV--- 215
Qy 298 GICINLIEVTW-----KSQKLYQPNNDYSEFMGNFSFMGVSVLMLFPVGNVI 349
Db 216 GRRLHLMRWVADATRVSKQVSL-----VLVTFPLAALT 252
Qy 350 RKFGH--VTGALVTPVWV-----LITGI-----VEPALVIFRNOASGL--V 386
Db 253 QRLGHPILGAFALGVILNSAPRTNRPLLDGVQITVAGLFAPVFPVLAGMVDVSQLRTP 312
Qy 387 AMFGTPLMLA-----VVGAIONLSKSTKXALFDSKEMAYIPLDOEKVWKAID 440
Db 313 AAMGTVALILTATVAKVVPALGARLG-----GLRGEALVAVGLNM-----KGGTD 361
Qy 441 VVARFGKSGALIQOGLLVICSGAMTPYIAVILFLIA-----IMVSAIKLKLFL 495
Db 362 LIVAIIVG-----VELGL-----SNEAYTMAVAVLVITASPLLILWE----- 401
Qy 496 AQSALKEQEVAAQEDSA 511
Db 402 -----KRAPPTQESBA 412

RESULT 13
US-09-252-991A-30725
Sequence 30725, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 30725
LENGTH: 534
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30725

Query Match 4.6%; Score 120.5; DB 4; Length 534;
Best Local Similarity 20.0%; Pred. No. 0.0034;
Matches 97; Conservative 71; Mismatches 145; Indels 171; Gaps 24;

Qy 63 IPFIKMLVPCALIFMLIVAKLSNLSKQALFVAVGTPPLIF--FALPPTIYV-----LR 117
Db 117

Db 110 LPARRFMNVVLCPLPAIPAFVSSFTWVSLSPMVEGLGAILVMTSKYPLVYLPVAATLR 169
Qy 118 DVLPHTPEADR-----LQAILP-----PGL-----GLVAILENMTF- 149
Db 170 GIDSLSEASAMGLSRQVFLRYTLPLRLPTLATGILVALHMLVERGALSTIRYQTF 229
Qy 150 AAFV-----VLAELMGSVLSLMF--WGFANEITKIHAKRFPALFGIGANISL 197
Db 230 TAIQEFELERSNATAMLSVLALCFLLM-----LELMR 267
Qy 198 ASGRAI--VWASKLRASVSEGVDPWGISLRLLAMNTVSGLV--VMAASYMMINKNVLTDP 253
Db 268 GRGLVHTGQSARABRVRVRFQKPLQALLATLVLTGSIPLVMGLW----- 317
Qy 254 RFTYPEEMOKKAKPKPMKMSFLYIDRSFYILLTLVIANG--ICINLIEVTWK 309
Db 318 -----LYEGSSASFPLMEIATST-----LISLSIARCGALLSCLLALPV--- 356
Qy 310 SQLKLYQPNNDYSEFMGNFSFMGVSVLIMLFVGNVIRKFGMLGALVTPVWVLLTG 369
Db 357 SILVVRTP-----GALARMT-----QRLPYLQAL--PGLVIALS 389
Qy 370 IVEPALVIFRNOASGLVAMFGTPLMLAVVGAIONLSKSTKXALFDSKEMAYIPLDQ 429
Db 390 LVYFSL-----HYLDATYOTGLLVA-----YALL-----FMPRLQ 421
Qy 430 E---QVKGKAIDVVAARFGKSG--GALLQOGLLVICSGAMTPYIAVILFLIAITWL 484
Db 422 APIRVALEKASPOQEBAARTIGTPLMAFRITLPIIPAIGA-----GFLVLFDTWKE 476
Qy 485 VSAT 488
Db 477 LTAT 480

RESULT 14
US-09-252-991A-32764
Sequence 32764, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 32764
LENGTH: 1162
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32764

Query Match 4.6%; Score 120.5; DB 4; Length 1162;
Best Local Similarity 21.1%; Pred. No. 0.0011;
Matches 81; Conservative 65; Mismatches 136; Indels 101; Gaps 19;

Qy 71 VVPCALIFMLIVAKLSNLSKQALFYAVG--TPPL-----IFPALPPTIYPLRDVILHP 122
Db 688 VIGIAVAV-LIGYAAAMWLIR-LIFSSPGDERPSELIMIVGILFTALPLVLF----- 737
Qy 123 TEFADRLQALIPPLLGLVAIILNMTFAAFVYLAELMGSVLSLMFGFANEITKIHAK 182
Db 738 -----VACGCGIYITSLSKLDRLIDITLYL--LILM--FYIEAAVFR--- 774
Qy 183 RFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISLRLLAMNTVSGLV-----LM 239
Db 775 -----GLGVAAARLAVARAL--AKRQNAKEGVDSFTVNEBPTLGIHQINQGSRLIR 825

```

295 IAGICINLIETWKSQLEQYPNNDYSEFMGIFSFMTGIVSVLLIMFVGGNVIRKFGW 354
      : ||      | . : : : :      : : : : : : : : : : : : : : :

```

```

01  | | | | |
02  | | | | |
03  | | | | |
04  | | | | |
05  | | | | |
06  | | | | |
07  | | | | |
08  | | | | |
09  | | | | |
10  | | | | |
11  | | | | |
12  | | | | |
13  | | | | |
14  | | | | |
15  | | | | |
16  | | | | |
17  | | | | |
18  | | | | |
19  | | | | |
20  | | | | |
21  | | | | |
22  | | | | |
23  | | | | |
24  | | | | |
25  | | | | |
26  | | | | |
27  | | | | |
28  | | | | |
29  | | | | |
30  | | | | |
31  | | | | |
32  | | | | |
33  | | | | |
34  | | | | |
35  | | | | |
36  | | | | |
37  | | | | |
38  | | | | |
39  | | | | |
40  | | | | |
41  | | | | |
42  | | | | |
43  | | | | |
44  | | | | |
45  | | | | |
46  | | | | |
47  | | | | |
48  | | | | |
49  | | | | |
50  | | | | |
51  | | | | |
52  | | | | |
53  | | | | |
54  | | | | |
55  | | | | |
56  | | | | |
57  | | | | |
58  | | | | |
59  | | | | |
60  | | | | |
61  | | | | |
62  | | | | |
63  | | | | |
64  | | | | |
65  | | | | |
66  | | | | |
67  | | | | |
68  | | | | |
69  | | | | |
70  | | | | |
71  | | | | |
72  | | | | |
73  | | | | |
74  | | | | |
75  | | | | |
76  | | | | |
77  | | | | |
78  | | | | |
79  | | | | |
80  | | | | |
81  | | | | |
82  | | | | |
83  | | | | |
84  | | | | |
85  | | | | |
86  | | | | |
87  | | | | |
88  | | | | |
89  | | | | |
90  | | | | |
91  | | | | |
92  | | | | |
93  | | | | |
94  | | | | |
95  | | | | |
96  | | | | |
97  | | | | |
98  | | | | |
99  | | | | |
100 | | | | |

```

QY 482 ----IMVSAT-----KINKFLAOSA 499
Db 875 IRLMIFVGAALGIIVLKLILDLSSQA 902

RESULT 17

US-09-107-532A-6563
Sequence 6563, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Arinello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 6563:

SEQUENCE CHARACTERISTICS:

LENGTH: 408 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...408

SEQUENCE DESCRIPTION: SEQ ID NO: 6563:

US-09-107-532A-6563

Query Match

Best Local Similarity 4.4%; Score 115.5; DB 4; Length 408;

Matches 88; Conservative 70; Mismatches 162; Indels 113; Gaps 16;

Db 102 FLIFPALPTVIYPRDVLHPTPEFADRLQALIPGLGLVAIL-----RMT 148

QY 27 YIIIVLAVIVAVIADTIIH---ASLSQGRASGIYITGLIARLFMGKQLEFGKRT 82

Db 149 F---AAFYVLAELMSVMSLMSFWGFRANETIKIHEAKRFYALFGIGANISLLASGRAIYW 205

QY 83 LRGALFY-----LLSTLALVLTPTIGALY-LVRFLNGRGY---TVSTATNAIYT 129

Db 206 ASKLASVSEGVDPWGISRLMLAMTIYSGVLVMSYWMINKVLTDPFRFVPEEKQK 265

QY 130 AVIPASKKEGGINYGLSTSLAAAVPFLGMIL----- 162

Db 266 KSAKPRMNKDSFLYIDRSPIYLLLLLVIAVGICINLIEVTWKQLKLYPNMNDYSR- 324

Db 163 -----LMMTD-FRFLIMPSIVLVEFVITLACVPEVKNIPSEEEKISLRNNINSFLEK 215

QY 325 ---FMGNFSFWTGV--VSYLIMLFGVAVIR-----KFGMLTGAIVTPMVLITGVF-- 372

Db 216 KALFITGIAFLMGLSYSSVLSFSSYTKVILHVAAGSFFVVAVLVITRPLTGRIPIDV 275

QY 373 -----FALVIFRNQASGLVAMPFTTPLMLAVVGAIONILSKSTKYALPSTKEMAYI 425

Db 276 KGEQVMTPSYLF--LTRGLFLLSVTNSVTLLISGALVGL-----GVTFMNGOAVCL 328

QY 426 PLDOEKYKGAIDVVAARFGSGGALIQ-----QGLVIGSIGAMTPYLAIVL 476

Db 329 KIVREHRSIALSTYFIGLDLGLGVGYLMGWLKSLTSRGLYIAG-----VLP 378

QY 477 LFIIMLVSATK 489

Db 379 LICTALYLVNGRK 391

RESULT 18

US-09-252-991A-17104

Sequence 17104, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubinfeld et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 17104

LENGTH: 546

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-17104

Query Match

Best Local Similarity 4.4%; Score 115.5; DB 4; Length 546;

Matches 104; Conservative 77; Mismatches 178; Indels 247; Gaps 22;

QY 18 PIHTHELKKVLEPMFPCITFENYVLRDTKDTLIVGAPSGAEAIPIFKWL----- 70

Db 18 PIKTHQVRLNDI-----VFFGSTALLIVLTAALIVPDSAGALLGRIOAMLSHSFGMY 71

QY 71 --VPCAILPMLIVA-----KLSNLSKQALFYAVGTPLIFPALPTVIY-----P 115

Db 72 YMLAIGAVLIFVAWAFSFRFTLKLGEHEKEDFSYGAWAGLFFSSGIGISILLYPAASBP 131

QY 116 LRDVHPTPEFADRLQALIPGLGLVALLRMTFAFYVLAELMSVMSLMSFWGFRANETI 175

Db 132 IDHLVHP-----PEGVPSGPOARQ-----ALDUTFLHMGJ----- 162

QY 176 TKIHEAKRFYALFGIGAN-----ISLLASGRAIYWASKLASVSEGVDPWGISRL 226

Db 163 ---HGMAIYALVGLAVGFAYRHRQPLALSALYPIIGERVYKKAAGVADCPGIFVTL 218

QY 227 L-----MAVTYSGVLVMSYWMINKVLTDP 252

Db 219 LGLVTLNIGALQVSSGLEVLTGMPSKGTLLAVILMSLVATTLAAVSGV----- 268

QY 253 PRFVPEEKQKKAQKPRAMKDSFLYLDSPYLLLTLLVIAVGICINLIEVTWKSQL 312

Db 269 -----EKGIKRLSNLN-----IVLSSLLFLVLCGSTLEL-----L 300

QY 313 KLYPNMNDY-----SEFMGNFSF-----WTGVSVLIML 342

Db 301 NGFVNGLDYLIDGLVLTKTFDLYVYTGAGAGKSEEMLGLMTLFYMAWMTISMAFPVGMFLAR 360

```

QY      343 FVCGNVIRKFGMLTGLVTPVWMLLGVFEFALVIFRNQASGLVAMFGTLP----- 394
DB      361 ISRGSRVRE-----LVCGVLLIPDGFTLAMSVEGNSALDVMNHGATDLKALAEOP 413
QY      395 -----MLAVVGAIONILSKSTKYALFDSCTKEMAYIPLDOEKVKGAIDVY 442
DB      414 SMSIVLLLEHYPLSKVIG-----LSIPGCVLF-----LTPADSG-----SWV 452
QY      443 AARFGSSG-----ALIQGLVITCGSIGM-----TPYLAVL 476
DB      453 LAMLSRSGGELDEBDAPNMLRIIMSAVVTLVITGL-FAQFTMQTVVVLAGLFPASAVLI 511
QY      477 LPTIAI 482
DB      512 LPMFGL 517

```

RESULT 19

```

US-09-328-352-6961
; Sequence 6961, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6961
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6961

```

```

Query Match      4.3%; Score 114; DB 4; Length 414;
Best Local Similarity 22.9%; Pred. No. 0.0012;
Matches 81; Conservative 71; Mismatches 130; Indels 72; Gaps 20;

```

```

QY      163 MSLMFWGPFANEITKIEAKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPW-G 221
DB      109 IISMSPFLPASPVSMLFIAR---GLOGIATGLAVSAIGAAIIDFSKLGSLINSIAPMIG 165
QY      222 ISLRILMANTIVS-----GLVL-MASYWINKNVLTDPREYNPEMOGKKA-----KP 270
DB      166 MAVGIFLTCSTIIOFSAPHLOLVFELCFLLICELLIS---FLTPETAQR-RSGALASLRP 221
QY      271 KM---NMKDSPLYDRSPYILLTLTLVAVGICINLIEVTKWSQQLQYPMNNDYSBPM 326
DB      222 NMAIPQTSALIST--SPINILMMV---SGFFELM-----PSL--LAKIF 262
QY      327 GNFSFWTGVVSVLIMLVFGNVIRKFGMLTGLVTPVWMLLGVFEFALVIFRNQASGLV 386
DB      263 HTSSAWLNGIMFWALMSG-----VGILTKRSTFRILLTGTLISAI-----GAI 309
QY      387 AMFGTPLMLAVV--VGAIQNLISKSTKALPDSYSTEMAMVPLDQOKKGAIAIDVVA 444
DB      310 VFIALINLTMNAVVLFGSIIITGVFGT--AFMGAISSVMPPLPBER-----AGIMMA 360
QY      445 RRGSGGALIQGLLV--ICGSIGAMT---PYLA-VILLFIINIVASATKLNK 492
DB      361 FVEASTLASIRAILAGVGVKIGLSTANSYSIFILLSLVALLMTIKKFKK 414

```

RESULT 20

```

US-09-134-001C-4637
; Sequence 4637, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

```

```

; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4637
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4637

```

```

Query Match      4.3%; Score 113; DB 4; Length 478;
Best Local Similarity 19.5%; Pred. No. 0.0018;
Matches 94; Conservative 75; Mismatches 174; Indels 138; Gaps 20;

```

```

QY      11 KLRSEFLMPIHT---HELRKVLV-----MFLMF-----CITFNVTY----- 43
DB      12 KYRKELIPVIGLLIITMALPIKEDALNDQAMFMFALPSTIIACITQPMITAVSIGRT 71
QY      44 -----LRDTKDTLIVAGPAGGAEAI-----PIKFWLVPCAIIMLYAK---L 85
DB      72 IMILVIGVDTK-TAVQGENSSIWLIAMAFISRGFVTKGLRRITALQFKLGKKTGL 130
QY      86 SNILSKQALFYAVGTPFLIFFA---LPTVYIPLADVL--HPTFPADRLQALPPGLGL 140
DB      131 AVSLGVVDLILAPAPPSNTARAGIIMP-IKSLSESGSSPRDSE----- 177
QY      141 VALINMTFAFYVLAELMGSVMLSLMFWGPFANEITKIEAKRFYALFG--IGANISLTA 198
DB      178 -----KMGALFITEPQGNLITSAF-----LTAMGNPFAQSLAEKT 215
QY      199 SGRAIYVASKLRASVSEGVDPWGISLRILMANTIVSGLVASWYWNKINVTLPREFNP 258
DB      216 AHVQITMMWMEVAAIIPG-----LISLIVPFIY-----KLYPP 250
QY      259 --EENQKGGKAKPKMNKDSFLYDR---SPYILLTLVAVGICINLIEVTKWSQQL 313
DB      251 TYKETPNAKWATPEOLEENGHSIAEKLVGVFTALALMWLGSTINVDATITAFIALAL 310
QY      314 IQYPMNNDYSBPMGNFSFWTGVV--SVLIMLVFGNVIRKFGMLTGLVTP-----V 363
DB      311 LLLTGLVAMSDILNLTGAMNTLWVFSVLVMAEQNLKGFIPWLSKLAQGLNGRSPV 370
QY      364 WLLTGVFEFALVIFRNQASGLVAMF-----GTPPLMAYVGAIONILSKSTKY 413
DB      371 LVLLILFEYSHYLFASATAHVASAMYAALLGVAVASGAPPLFSALMLGFFGNLSTHY 430
QY      414 A 414
DB      431 S 431

```

RESULT 21

```

US-09-252-991A-18288
; Sequence 18288, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18288
; LENGTH: 578

```



```

; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18288

```

Query Match 4.3%; Score 113; DB 4; Length 578;
Best Local Similarity 22.4%; Pred. No. 0.0025;
Matches 120; Conservative 65; Mismatches 153; Indels 198; Gaps 33;

```

QY      1 MTKEHEKPGKRSFLMPFIHTHELKVLMPMLMFPCITNNYATLRDQOTLLYGARGSGA 60
Db      170 MHRSGARP--RLRS-----RRLAVLLPILL--AAAGFTLL--TTEVVG----- 210
QY      61 EALPFIKEFWLVPICAIIFMLIYAKLSNIISKOLFEYAVGTPELFIFFALPEFTVYIPUDVL 120
Db      211 -----LPLALAADLOVSAQAGLLVSLFASFVAAGPF----- 243
QY      121 HPTFEADRLQAILPPEGLGL-----VALIRMTFAAFYVALIELWGSV-----ML 164
Db      244 -----LTPAALGVERKRLPEVACL--LFPAAANAALAVAGDITMAVARFVPAL 289
QY      165 SL-MWGFANETITTKIHEA-----KRPYAL--FGI-GANISLASGAIYMASTKLARSVSE 215
Db      290 ALPFWMAVASE-TAAHLGSPREGRVAVLFPQIYAAVTLGIPIGILL----- 336
QY      216 GVDPMG--ISRLTAMMTIVSGLVMAVYMWINKOVLTPDPRFYNPBEMQKKKGAKPKM 273
Db      337 -ADAWGRLAPALALALAKALLLAA--WL-----PRI-----PGRPGIS 374
QY      274 MKDSFLYADRSPYI---LTLTLVLVAYAGICIMLEVTWSQLKQYPMNDYSE----FM 326
Db      375 IRSQASVYL-RQPLVUGHLLSLIVF-----TGMFTPTTYLADIIQLRLAGFS 419
QY      327 GNFSFWT---GVSVLIMLFVGVGNIRKFGMLTGALV-----TPVWLLTGIVFAL 375
Db      420 GSLVGMTLMGFAGVGL-----GN-----WLGRLVDRSPGLATLLFVLNALGMLAL 467
QY      376 VIFRQQA---SGLVAMFGTTPMLAVVGAIQNIILSKTKYALFDSKTEMAYIPLDQEQK 432
Db      468 VPTLGNAWMLLATLAIWGAQAL--FIVGOVR-VMSARAPAF----- 509
QY      433 VKKGAIDIVVAPKSGGAL-----IQGGLVIGSGISAMT--PYLAVITLFI 481
Db      510 ---AASLIVISAGNAGIGIALGSNVYIDSGSLAPLGEVAGLVCILAAMAAVALLMLLA 562

```

```

RESULT 22
US-09-112-096-15
Sequence 15, Application US/09112096
Patent No. 6194152
GENERAL INFORMATION:
APPLICANT: Reinher Laus
APPLICANT: Michael H. Shapiro
APPLICANT: Larissa Tsavaler
TITLE OF INVENTION: Prostate Tumor Polynucleotide and
TITLE OF INVENTION: Antigen Compositions
FILE REFERENCE: 7636-0015.30
CURRENT APPLICATION NUMBER: US/09/112, 096
CURRENT FILING DATE: 1998-07-09
EARLIER APPLICATION NUMBER: 60/056, 110
EARLIER FILING DATE: 1997-08-20
NUMBER OF SEQ ID NOS: 29
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 15
LENGTH: 1095
TYPE: PR1
ORGANISM: Homo sapiens
US-09-112-096-15

```

Query Match 4.3%; Score 113; DB 3; Length 1095;
 Best Local Similarity 21.0%; Pred. NO. 0.0064;
 Matches 85; Conservative 54; Mismatches 121; Indels 144; Gaps 19;
 43 VLRLTKDLYLGAPSSGAEALPIRFYFWLVPCALIFMI-----YAKLSNLSKQALFY 96

```

Db 677 ISRDITK-----|::|||::|::|
QY 97 AVG---TFPLI-----FFALFPTV-IYLRDLVLPTEFADRLQAILPGLGLVA 142
   ||::||::|
Db 718 YVAFPTSPFVFSNMVVFYIAFLLFAVYLLMDHSPHPPELV-----LVSIVF 767
   ||::||::|
QY 143 IL-----BNMTFAAFYVLAELMGSVLSI-MWGFANEITIKIHEAKRYALFGIGANISL 196
   ||::||::|
Db 768 VLFCEVAVQWVYNGVNYFTDLN-NMMDTLGLGYFLAGIVFRLL-----SSNKSS 81.5
   ||::||::|
QY 197 LASGRAIYMAKSLRASVSEGVDPWGISLRLLMAMTIVSGLVIMASYMWINKVLTDPREY 256
   ||::||::|
Db 816 LYSGRVIFC-----LDYILFTLRLLIHIFV-----84.0
   ||::||::|
QY 257 NPEEMOKKKGAKPPKMNKDSFLYIDRSEYIILLVLVAYGICMLI-----304.0
   ||::||::|
Db 841 -----SRNLGPICIMLQRMIL-IDVFFFLFLFVAMVVAFCVARGOIGLRONEQWRMIF 89.2
   ||::||::|
QY 305 -EYVMKSKLKL--QYPMNDYSEFMGNSEFMWGVVSVLIMLVGGAVIKKF-CWMLGALV 36.0
   ||::||::|
Db 892 RSVIYEPLAMFGQVSPVDGTTTYPACTFTGNESSKPLCVBLDEHNLPRFPEMITIPLV 95.1
   ||::||::|
QY 361 TPVWTLTGIVFPFALVIFRNQASGLVANEGTTPMLAVVGAIG 404
   ||::||::|
Db 952 C-IYMLSTNILLVNL-----LVANFSGY-----VGTQV 978
   ||::||::|

```

```

RESULT 23
US-09-328-352-4974
; Sequence 4974, Application US/09328352
; Patent No. 6562358
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4974
; LENGTH: 419
; TYPE: prt
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4974

```

| | | | | | | |
|-----------------------|--------|--|-------|-----------------|-------------|---------|
| Query Match | 4.3%; | Score 112.5; | DB 4; | Length 419; | | |
| Best Local Similarity | 20.0%; | Pred. No. 0.0017; | | | | |
| Matches | 94; | Conservative | 61; | Mismatches 148; | Indels 167; | Gaps 22 |
| Qy | 84 | KLNLKSKOAL-----FYAVGTPELLIPFALPEPTVYTPRLDYLHPTFADRLQAILPP | 135 | | | |
| Db | 22 | KLNLKSTVTAFLALAIAGAFALGT-----TEFS----- | PM 52 | | | |
| Qy | 136 | GLLGLVA-----ILRWTFAPAFVYLAELKWSVNLSTMFQFANE----- | 174 | | | |
| Db | 53 | GLLENLIANDLGISIPTAGMLITGYALGYMLGAPFMTLFGGFARRNALIFLMAIPTVGNL | 112 | | | |
| Qy | 175 | -----TTKIHEARFPYALPGIGINISLLASGRATVWASKLRASVE----- | 215 | | | |
| Db | 113 | IAAFSPNYSMLGARLLITSLNHG-----AFFGISGVA-----ASTVPHKQASAAVATWFM | 163 | | | |
| Qy | 216 | -----GVD--PW-----GISRLILMAMTYISGIVLMAWYVINKVNTDPRFYPEEMQ | 262 | | | |
| Db | 164 | GLTIANIGGVPLATWVGONIGWRISFLAISTGLITMLALM-----KALP | 208 | | | |
| Qy | 263 | KKKKGAAPKMMKDSPLYLDRSPYLLITLLVIAYGICINLIEVTWSQLKQVPPNNDY | 322 | | | |
| Db | 209 | QGWVAQRP--NVRKAEKLVLTETPVYALLFTVLGAGAMFTLYYI-----APSLTEF | 258 | | | |
| Qy | 323 | SEFPNGNSFPTQGVSVLIMLFVGGNVIRKRPMLTGALVTPMVULLTGIVFPALVIFRQA | 382 | | | |
| Db | 259 | THASPTF-----ITFMVLIGVGSII-----GNHLGGRPADISINKTKIIGFVVL----- | 302 | | | |

Qy 383 SGVAMECTPLMAVVG-AIONLSKSTKALFDSTKEMAYPLDQ-----QKVGK 436
Db 303 --LIVMTPTFPILAQSOIGAAIALVIGCAATFALP-----PLQNRVNSVAHEADGL 352
Qy 437 A-AIDVVARFGKSGGALIOGGLV-----ICSGIAMPYAVILL 477
Db 353 ASSVNIQAFNIGNAVGA--AGALVLDLQMGKSVSPFAGALLAGGLLV 400

RESULT 24

US-08-597-236-11
; Sequence 11, Application US/08597236
; Patent No. 5733765
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesea
; APPLICANT: MOULET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; NUMBER OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/597,236
; FILING DATE:
; CLASSIFICATION: 426
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-597-236-11

Query Match 4.2%; Score 111; DB 1; Length 360;
Best Local Similarity 17.2%; Pred. No. 0.002;
Matches 74; Conservative 80; Mismatches 142; Indels 134; Gaps 19;
Qy 103 LIFPALFPTV--IYPLRDVL-HPTFADRLQAILPPGLGLVAIRNMTFAFVYLAELM 159
Db 2 VIFFLFPMIMAMITMLTLRQKQIOKTIKCVLTFGLGFSASRASSVGTDTLYE-- 59
Qy 160 GSVMLSMF-----WGFANEITKIHAKRFYALFGIGANISLASGRAIVAMSKLRAS 212
Db 60 -NIFKSINYGISAENNGY-----VIYNKLGISVFGY-----TGEHITANSVLIT 104
Qy 213 VSEGVDPNGISRLIMA-----MTIVSGIYLMASVWINKVLTDP 254
Db 105 ILIGFIWKVAEHVATFLYISLFYATSFNISROFIAMGLVVAISFALDKKVM--- 160
Qy 255 FYNPEEMQKKGAKPKNMKDSFLYDRSPYILLTLTLVAYGICINLIEVTKSQKL 314
Db 161 -----PFFILTVALTLPHATAIYAFVYVLT--KV 188
Qy 315 QYPMNDYSEF---MGNFSEFWTGVSVYLIMLVGAGNVIKRGWLGT-----A 358
Db 189 HMDVKKTLISFPIITFAFIF-DALINIFVRFPPHYEM-----YITGQFNISDQGRV 242

Qy 359 LVTEVMVLNGIVPEFALVIFRNQASGLV-----AMPFTPLMLAVVGAIONLSKSTKY 413
Db 243 VLVKIFILL--LIFLFLFKYSYALISBOHSLALITVGLSIGVFNIL--LNRI 297
Qy 414 ALFSTKEMAYPLDQCKVKAIDVVARFGKSGA--LIOGGLVI-----C 462
Db 298 EMFYSILSIYFIP-----ALDYISLKFQKQAVRMLTIGLILLVLPYIIGVS 347
Qy 463 GSIAMTPYL 472
Db 348 GNYSGILPYV 357

RESULT 25

US-08-746-682A-11
; Sequence 11, Application US/08746682A
; Patent No. 5786184
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesea
; APPLICANT: MOULET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; NUMBER OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,682A
; FILING DATE: 14-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597,236
; FILING DATE: 20-JUN-1995
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-746-682A-11

Query Match 4.2%; Score 111; DB 1; Length 360;
Best Local Similarity 17.2%; Pred. No. 0.002;
Matches 74; Conservative 80; Mismatches 142; Indels 134; Gaps 19;
Qy 103 LIFPALFPTV--IYPLRDVL-HPTFADRLQAILPPGLGLVAIRNMTFAFVYLAELM 159
Db 2 VIFFLFPMIMAMITMLTLRQKQIOKTIKCVLTFGLGFSASRASSVGTDTLYE-- 59
Qy 160 GSVMLSMF-----WGFANEITKIHAKRFYALFGIGANISLASGRAIVAMSKLRAS 212
Db 60 -NIFKSINYGISAENNGY-----VIYNKLGISVFGY-----TGEHITANSVLIT 104
Qy 213 VSEGVDPNGISRLIMA-----MTIVSGIYLMASVWINKVLTDP 254
Db 105 ILIGFIWKVAEHVATFLYISLFYATSFNISROFIAMGLVVAISFALDKKVM--- 160
Qy 255 FYNPEEMQKKGAKPKNMKDSFLYDRSPYILLTLTLVAYGICINLIEVTKSQKL 314

```

Db 161 -----PMFLITVATLPHATPAVFPVWLT-KV 188
Qy 315 QYPMNDYSEF---MGNFSTWGVSVLMLPGGVNIRKFGMLTG-----A 358
Db 189 HMDVKTLISIPITIFASFIF-DAILNIFVRFPPHYEM-----YITQFNISDQGGRV 242
Qy 359 LVTPVWMLTGTIVFPAVIFPNQASGV-----AMFGTPLMLAVVGAONILSKSTKY 413
Db 243 VLVKIFILL---ILFTLFLFYKSYALISECHOSIALITVGLSIFVFNHIL--LNRI 297
Qy 414 ALPSTKEMAYIPLDOBKVKAIDVVAARFGSGA--LIQGLVLI-----C 462
Db 298 EMFSTISIVTPI-----AIDYISLKFKQKQAVRLMLTIGILLITLVPIYIQVS 347
Qy 463 GSTGAMTPYL 472
Db 348 GNYSGILPEYV 357

```

RESULT 26

```

US-09-252-991A-25267
; Sequence 25267, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25267
; LENGTH: 770
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25267

```

```

Query Match 4.2%; Score 111; DB 4; Length 770;
Best Local Similarity 20.0%; Pred. No. 0.0062;
Matches 83; Conservative 62; Mismatches 123; Indels 146; Gaps 23;

```

```

Qy 67 KFWLVPCALI-FMLIYAKLSN-----ILSKQALFYAVGTPFLI--FPALFPT 111
Db 345 RMYVIMPIVFTYSLAYIDRANYGAAAGMADLRITPGSLGLAFIFGFFQVPG 404
Qy 112 VIYPLRDVLAHTEPADRLQAILPPG--LGLVA--ILRWTPAIFYVLAELMGVMS 165
Db 405 AIVAEKRSVKLIF---VSLIMGGLATLTGMVANVYLLIGIRFLGVAEAAVAMPMLVY 461
Qy 166 LMFPGFANEITKIHAKRFVLPFGANISL--IASGRVIAWSTKRAVSQVDP--PWGIS 223
Db 462 LCHN---FTRARRSR-----ANTFLILGPNVTILMWSVVGVLVHFSGRW--- 504
Qy 224 LRLIMATVIVGL--VIMASYW-----INKVLTDPFVYPEEMQKQK 266
Db 505 -----MFLIEGLPAVIMAFIMWRLVDRPRQAAWLSSEKRALED-----ALAAEQ 551
Qy 267 GAKRPMNKOSFLYLDGSPYLLITL-----LVIAVGICINILEVTWK 309
Db 552 GIKFVKVYREAF---RSPKVIIVLSQYFWSIGYGFVLMPLSLKKGANIDIIEAGWL 607
Qy 310 SQL-----TLQ-----YP-----NMNDYSEF-MGNFSTWGVSVYLI 340
Db 608 SALPYLAAVIMLVGWSASDRLOQRKRFPVPPLLIASIAFYGYSALGSEHFW---LSTAL 664
Qy 341 MLFVGAVIRKFGMLTGVLPVWMLTGTIVFPAV---VIFPNQASGLVAMFGT 391
Db 665 LVLAGACMVAVYG-----PFPAIVPEVLPAVYAGGAMALINS 701

```

RESULT 27

```

US-09-107-532A-6105
; Sequence 6105, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

```

```

NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6105:
SEQUENCE CHARACTERISTICS:
LENGTH: 478 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...478
SEQUENCE DESCRIPTION: SEQ ID NO: 6105:
US-09-107-532A-6105

```

```

Query Match 4.2%; Score 110; DB 4; Length 478;
Best Local Similarity 21.3%; Pred. No. 0.0039;
Matches 108; Conservative 77; Mismatches 195; Indels 126; Gaps 25;

```

```

Qy 27 VLPWFMEFCTTENVYTLVRLDKTLIVGAPSGAEALPFIK---FW-----LV 71
Db 19 LLSLFPFPEF-----NHQNKKSQVMAOKTTATVABEKKKKGFWSKITGVSLPLVLI 72
Qy 72 VPCALIFMLIYAKI-SNLSKQA-----LFYAVGTPFLI-----FPALFPYIY 114
Db 73 MVVLILITAMALGKPLTNMIGRIALVIFGNLFPHIGTKIPVKSYLEGGSVFALFASA-- 130
Qy 115 PLRDVLAHTEPADRLQAILPPGILGLVAI-----LKNWTPAIFYVLAELMGVMS 163
Db 131 -----ALATFGIILPEYVAVKSTENFVNMGFWMDYIALILITGSLIGMARN 174
Qy 164 -----LSLMFGFANEITKIHAKRFVA-----LFGIG-AN-----ISLAS---GRAIV 204
Db 175 LMKASVRF-----IPVALLISWVSGFFAVGLVGMILGNGFANSVLYISLPMAAGVGAGAV 230

```


QY 179 -----HEAKRFYALFGIGANISL-LASGRAIVWASKLRASVS-----EGVDP 219
 Db 137 LVIALAKRELMNDRAAAAAAIIIGGATWALGIISSAAQIQANKTSLEPSTYNTLTGVP 196
 QY 220 W--GISLRLLMAMTIVSGVLMA-SYWMINK--NVLT-----DPRFNPPEMOGKKGAKP 270
 Db 197 FTEETFLMOSIMNTIIVLSIAIVWASAPKNSVKTIDSPVOF--BEETKNWAKSTRP 254
 QY 271 KKNMDSFLYDRSPYLLLTLLVAVGICINLIEVTWKSQKQYPMNDYSEFMGNS 330
 Db 255 GD-----WLENSP---LTIITIVVGLIMWFEERS-KSNPIIAISSINTY-----NEV 298
 QY 331 F 331
 Db 299 F 299

RESULT 30

US-09-328-352-6605
 ; Sequence 6605 Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: CTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 6605
 ; LENGTH: 548
 ; TYPE: PRP
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-6605

Query Match 4.1%; Score 108.5; DB 4; Length 548;
 Best Local Similarity 22.0%; Pred. No. 0.007;

Matches 105; Conservative 58; Mismatches 140; Indels 175; Gaps 27;

QY 99 GTPLIFPALFPTVYPLRDVLP--TEPADRLQAILPGLGLVAILRNWTFAPYVLA 157
 Db 80 GTTNVVFSLFPAIIIFLTGI--PLAVYARVN-----IDLDIRGAGG--YI--- 125
 QY 158 LMGSVMLSMFEGFANETKTHAKRFYALFGIGANISLASGRAI---VWASKLRAS 212
 Db 126 --GSVLTSLIFASF---TFI-----FPALIEG---SMAQGLLGLGIPLMAGYLIST 169
 QY 213 VSEGVDP--WGI-SLRLLMAMTIVSGVLMA--ASYWMINKNVLTDPFNPPEMOGKK 265
 Db 170 VM--VILPVIYGMKALSKIQVWTTPLMLVLMIGPVAYL-----IYOEPLVVSQFATFTGH 222
 QY 266 KGAKRPMNMKDSFLYDRSPYLLLTLLVAVGICINLIEVTWKSQKQYPMNDYSEF 325
 Db 223 EGFAP-----VDMAALMGAGICISLI-----MQIGEQDIYARF 256
 QY 326 M-----GNFSFWTGVSV-----LIMFVGSNVTRKF----- 352
 Db 257 MPAKTKENSKAWMAAVISAGPGWILGAIKQIIGAFGLFGLTKIPGVNSTEPVOQFNA 316
 QY 353 -----GMLTGALVTPVAVL-----LTGIYFPALVIFR 379
 Db 317 FHDMLPGMAALTILAVILVVISQIKINVTNAYSGLAMTSAYTRISKHPRGVIVFIV--- 373
 QY 380 NQASGLVAMFGTTPMLAVVVG-----AIQNILSKST-----KVALFDSTKEMAYIPLDQ 429
 Db 374 NLAIALALMEGMPFAVLGKILGFYSNFAIAMVVAVATDISIKYVLKLSKPEPR-RDM 432
 QY 430 EOKVK--GKAIDV-----VAARFKSGGALIQGGLVYCGSIGANTPYLAVILFIT 480
 Db 433 LYNVAVGVAVLVSAGLSIAAFGLIGSFL-----ADYSPILAVLAVFL 478

RESULT 31

US-09-555-313B-2
 ; Sequence 2, Application US/09555313B
 ; Patent No. 6506580
 ; GENERAL INFORMATION:
 ; APPLICANT: PICHMEISTER, Rudolph et al.
 ; TITLE OF INVENTION: Splicing variants of the human serotoninergic receptor
 ; FILE REFERENCE: P06762US00/BNS
 ; CURRENT APPLICATION NUMBER: US/09/555,313B
 ; PRIOR FILING DATE: 2002-08-13
 ; PRIOR APPLICATION NUMBER: FR 97/15037
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 760
 ; TYPE: PRP
 ; ORGANISM: Homo sapiens
 US-09-555-313B-2

Query Match 4.1%; Score 108.5; DB 4; Length 760;
 Best Local Similarity 21.8%; Pred. No. 0.011;

Matches 62; Conservative 54; Mismatches 93; Indels 75; Gaps 15;

QY 269 KPKMMKDSFLYDRS-----PYILLTL--VIANGICINL--IEVTWKSQKQYPM 319
 Db 378 KPSMDKLDANVSSEEGSVKRVLLTFLSTVILMAIIIGNLVAVAVCMRQLRIKTY 437
 QY 320 NDYSEFMGNFSFWTGVSVLIMLFPVGNVIRKFGMLTGALV---TPVMVLLTGIVFPAL 375
 Db 438 -----FIVSLAFADLDVSVLWMPFGLIELVQDI-WYGEVCLVRTSIDVLTTSITHL 491
 QY 376 -----VIFRNASGLVAMFGTTPMLAVVVG--AIQNILSKSTYALFD 417
 Db 492 CCISLDRYAICCCPLVYRNK-----MTPLRIALMLGCGWVIFPFI----- 533
 QY 418 STKEAAYIPLDQEQVKVKKALIDVYAR-FKSGGA-----LIQGLVIVGSGIGAMTPY 471
 Db 534 -----FLPIQGMNNDI--IIDLEKRRQNNSNTYCVPMNRPVATCSVAPYIFP 585
 QY 472 LAVILFPIAIWVSATLKLFLAQSALKEQEVNQEASAPASS 515
 Db 586 LLMVLAV-YRIYVTRAKHAHQIQMLQRA-----GASSSESRFQSA 623

RESULT 32

US-09-252-991A-24105
 ; Sequence 24105 Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 24105
 ; LENGTH: 816
 ; TYPE: PRP
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-24105

Query Match 4.1%; Score 108.5; DB 4; Length 816;
 Best Local Similarity 19.3%; Pred. No. 0.013;
 Matches 57; Conservative 56; Mismatches 99; Indels 83; Gaps 10;

QY 134 PPGLLGLVAILRNWTFAPYVLAELMGSVMLSMFEGFANETKTHAKRFYALFGIGAN 193


```

Db      20 VLLTFLSTVILMALIGLWVAVCMDRQLRKIKTNV-----FIVSLAFADLVSVLVM 74
Qy      342 LFVGNVTRKRGWLTGALV---TPVMVLTGIVFPAL-----VIFRQQA 382
Db      75 PFGAIEIVQDI-WIYGEVFCIVRTSLDVLTTASIFHLCCISLDRYVAICCPVLYRNK- 132
Qy      383 SGLVAMFGTTPMLAVVVG---AIGNILSKSTKYALFDSTKEMAYIPLDQEQVKYKAAI 439
Db      133 -----MTPLRIALMGCCWVILPTFIS-----FLPIMQGNNIG--II 167
Qy      440 DVNAAR-FGKSGA-----LIIQGLLVICGSIGAMTPYLAIVLLEFIITAIWLSATKLNKL 493
Db      168 DLIERKRNQNSNSTYCVFMVNRKPYAITCSVVAFYIPFLMWLAY-YRIYVTAKEHAHQI 226
Qy      494 FLAQSALKEQVVAQEDSAPASS 515
Db      227 QMLQRA-----GASSESRPOSA 243

```

RESULT 38

```

US-09-555-313B-24
; Sequence 24, Application US/09555313B
; Patent No. 6506580
; GENERAL INFORMATION:
; APPLICANT: FISCHMEISTER, Rudolph et al.
; TITLE OF INVENTION: Slicing variants of the human serotonergic receptor
; FILE REFERENCE: P067620S00/BAS
; CURRENT APPLICATION NUMBER: US/09/555,313B
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: FR 97/15037
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 24
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-555-313B-24

```

```

Query Match      4.0%; Score 105.5; DB 4; Length 380;
Best Local Similarity 22.1%; Pred. No. 0.0085;
Matches 58; Conservative 50; Mismatches 83; Indels 71; Gaps 14;

```

```

Qy      287 ILLTLL--VIAVIGICNL---IEVTWKSQKLQYPNNDYSEFMGNSFWTGVSVLIM 341
Db      20 VLLTFLSTVILMALIGLWVAVCMDRQLRKIKTNV-----FIVSLAFADLVSVLVM 74
Qy      342 LFVGNVTRKRGWLTGALV---TPVMVLTGIVFPAL-----VIFRQQA 382
Db      75 PFGAIEIVQDI-WIYGEVFCIVRTSLDVLTTASIFHLCCISLDRYVAICCPVLYRNK- 132
Qy      383 SGLVAMFGTTPMLAVVVG---AIGNILSKSTKYALFDSTKEMAYIPLDQEQVKYKAAI 439
Db      133 -----MTPLRIALMGCCWVILPTFIS-----FLPIMQGNNIG--II 167
Qy      440 DVNAAR-FGKSGA-----LIIQGLLVICGSIGAMTPYLAIVLLEFIITAIWLSATKLNKL 493
Db      168 DLIERKRNQNSNSTYCVFMVNRKPYAITCSVVAFYIPFLMWLAY-YRIYVTAKEHAHQI 226
Qy      494 FLAQSALKEQVVAQEDSAPASS 515
Db      227 QMLQRA-----GASSESRPOSA 243

```

RESULT 39

```

US-08-446-822-8
; Sequence 8, Application US/08446822
; Patent No. 5766879
; GENERAL INFORMATION:
; APPLICANT: SYNAPTIC PHARMACEUTICAL CORPORATION
; TITLE OF INVENTION: DNA ENCODING 5-HT4 SEROTONIN RECEPTORS
; TITLE OF INVENTION: AND USES THEREOF

```

```

; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: COOPER & DUNHAM
; STREET: 30 ROCKEFELLER PLAZA
; CITY: NEW YORK
; STATE: NEW YORK
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,822
; FILING DATE: June 1, 1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: White, P., John
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 42667-A-PCT-US/JPM/MAT
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-446-822-8

```

```

Query Match      4.0%; Score 105.5; DB 1; Length 388;
Best Local Similarity 22.1%; Pred. No. 0.0087;
Matches 58; Conservative 50; Mismatches 83; Indels 71; Gaps 14;

```

```

Qy      287 ILLTLL--VIAVIGICNL---IEVTWKSQKLQYPNNDYSEFMGNSFWTGVSVLIM 341
Db      20 VLLTFLSTVILMALIGLWVAVCMDRQLRKIKTNV-----FIVSLAFADLVSVLVM 74
Qy      342 LFVGNVTRKRGWLTGALV---TPVMVLTGIVFPAL-----VIFRQQA 382
Db      75 PFGAIEIVQDI-WIYGEVFCIVRTSLDVLTTASIFHLCCISLDRYVAICCPVLYRNK- 132
Qy      383 SGLVAMFGTTPMLAVVVG---AIGNILSKSTKYALFDSTKEMAYIPLDQEQVKYKAAI 439
Db      133 -----MTPLRIALMGCCWVILPTFIS-----FLPIMQGNNIG--II 167
Qy      440 DVNAAR-FGKSGA-----LIIQGLLVICGSIGAMTPYLAIVLLEFIITAIWLSATKLNKL 493
Db      168 DLIERKRNQNSNSTYCVFMVNRKPYAITCSVVAFYIPFLMWLAY-YRIYVTAKEHAHQI 226
Qy      494 FLAQSALKEQVVAQEDSAPASS 515
Db      227 QMLQRA-----GASSESRPOSA 243

```

RESULT 40

```

US-09-328-314-8
; Sequence 8, Application US/09328314
; Patent No. 6331401
; GENERAL INFORMATION:
; APPLICANT: Gerald, Christophe
; APPLICANT: Hartig, Paul R.
; APPLICANT: Branchek, Theresa
; TITLE OF INVENTION: DNA Encoding 5-HT4 Serotonin Receptors And Uses Thereof
; FILE REFERENCE: 42667-AZ-PCT-US
; CURRENT APPLICATION NUMBER: US/09/328,314
; FILING DATE: 1998-04-03
; EARLIER APPLICATION NUMBER: 08/446,822
; EARLIER FILING DATE: 1995-07-31
; EARLIER APPLICATION NUMBER: PCT/US93/12586
; EARLIER FILING DATE: 1993-12-22

```


EARLIER APPLICATION NUMBER: 07/996,772
EARLIER FILING DATE: 1992-12-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 388
TYPE: PRF
ORGANISM: Homo sapiens
US-09-328-314-8

Query Match 4.0%; Score 105.5; DB 4; Length 388;
Best Local Similarity 22.1%; Pred. No. 0.0087;
Matches 58; Conservative 50; Mismatches 83; Indels 71; Gaps 14;

QY 287 ILLTL--VIAVCINL--IEVTKSOLKQYNNNDYSEFMGNSEFWTGVSVLIM 341
DB 20 VVLLFPLSTVIMAILGNLLMVAVCGMDROLRIKIKTNV----FIVSLAFADLVSVLVM 74
QY 342 LFGGVNIRKFGMLTGALV---TPVMVLLTGIVFPAL-----VIFRNOA 382
DB 75 PRGALBLVODI-WIYGEVPCLVRTSLDVLITASTFHLCCISLDRYAICQPLVYRNK- 132
QY 383 SGLVAMFGTTPMLAVVG--AIONILSKSTKYLFDSTKEMAYIPLDQOKVKGAAI 439
DB 133 -----MTPLRIALMLGGCWVLPTEIS-----FLPIWQGNNG--II 167
QY 440 DVVAAR-FKSGGA-----LIOGLVITGSGIGANTPYLAVILLPIAIMLVSATKNTL 493
DB 168 DLIERKRNQNSNSTYCVFMVKNRPYATCSVVAFYIPLLMVLAAY-YRIYVTAKEHAHQI 226
QY 494 FLAGALKEQEAQEDSAPASS 515
DB 227 QMLQRA-----GASSESRQSA 243

Search completed: November 25, 2003, 10:17:11
Job time : 26 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:16:18 ; Search time 54 Seconds

(without alignments)
1759.042 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 2630

Sequence: 1 MTKREKPFGLTSFLWPIH.....AQSLAKQEVAAQEDSAPASS 515

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 673684 seqs, 18443283 residues

Total number of hits satisfying chosen parameters: 673684

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubppaa/PCR_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubppaa/PCRTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubppaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubppaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubppaa/US10A_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-----------------------|--------------------|
| 1 | 2630 | 100.0 | 515 | 9 US-09-892-851-2 | Sequence 2, Appl1 |
| 2 | 162 | 6.2 | 538 | 10 US-09-738-626-3617 | Sequence 13617, Ap |
| 3 | 132.5 | 5.0 | 463 | 9 US-09-815-242-10144 | Sequence 10144, A |
| 4 | 129 | 4.9 | 462 | 9 US-09-815-242-10161 | Sequence 10161, A |
| 5 | 125 | 4.8 | 587 | 14 US-10-173-519-5 | Sequence 5, Appl1 |
| 6 | 125 | 4.8 | 587 | 15 US-10-167-994-12 | Sequence 12, Appl1 |
| 7 | 120.5 | 4.7 | 713 | 14 US-10-014-717-11 | Sequence 13892, A |
| 8 | 120.5 | 4.6 | 395 | 9 US-09-815-242-13892 | Sequence 11741, A |
| 9 | 120.5 | 4.6 | 465 | 9 US-09-815-242-12171 | Sequence 12171, A |
| 10 | 118 | 4.5 | 466 | 9 US-09-815-242-13831 | Sequence 13831, A |
| 11 | 117.5 | 4.5 | 430 | 9 US-09-815-242-5678 | Sequence 5678, Ap |
| 12 | 116.5 | 4.4 | 477 | 9 US-09-815-242-13840 | Sequence 13840, A |
| 13 | 115.5 | 4.4 | 410 | 10 US-09-738-626-3861 | Sequence 3861, Ap |
| 14 | 115 | 4.4 | 392 | 12 US-10-181-319-4 | Sequence 4, Appl1 |
| 15 | 115 | 4.4 | 708 | 10 US-09-738-626-5024 | Sequence 5024, Ap |

| | | | | | |
|----|-------|-----|------|------------------------|--------------------|
| 16 | 113 | 4.3 | 1095 | 9 US-09-759-143-778 | Sequence 778, App |
| 17 | 113 | 4.3 | 1095 | 9 US-09-780-669-778 | Sequence 778, App |
| 18 | 113 | 4.3 | 1095 | 9 US-09-822-827-778 | Sequence 778, App |
| 19 | 113 | 4.3 | 1095 | 10 US-09-895-793-778 | Sequence 778, App |
| 20 | 113 | 4.3 | 1095 | 10 US-09-895-814-778 | Sequence 778, App |
| 21 | 113 | 4.3 | 1095 | 12 US-10-144-678A-778 | Sequence 778, App |
| 22 | 113 | 4.3 | 1095 | 12 US-10-294-025-778 | Sequence 778, App |
| 23 | 113 | 4.3 | 1095 | 14 US-10-012-896-778 | Sequence 778, App |
| 24 | 113 | 4.3 | 1095 | 15 US-10-205-823-421 | Sequence 421, App |
| 25 | 113 | 4.3 | 1268 | 10 US-10-171-319-11 | Sequence 11, Appl1 |
| 26 | 113 | 4.3 | 1681 | 10 US-09-920-653-3 | Sequence 3, Appl1 |
| 27 | 111.5 | 4.2 | 1083 | 10 US-09-738-626-6914 | Sequence 6914, Ap |
| 28 | 111 | 4.2 | 741 | 10 US-09-738-626-4455 | Sequence 4455, Ap |
| 29 | 111 | 4.2 | 1095 | 9 US-09-759-143-780 | Sequence 780, App |
| 30 | 111 | 4.2 | 1095 | 9 US-09-780-669-780 | Sequence 780, App |
| 31 | 111 | 4.2 | 1095 | 9 US-09-822-827-780 | Sequence 780, App |
| 32 | 111 | 4.2 | 1095 | 10 US-09-895-793-780 | Sequence 780, App |
| 33 | 111 | 4.2 | 1095 | 10 US-09-895-814-780 | Sequence 780, App |
| 34 | 111 | 4.2 | 1095 | 12 US-10-144-678A-780 | Sequence 780, App |
| 35 | 111 | 4.2 | 1095 | 12 US-10-294-025-780 | Sequence 780, App |
| 36 | 111 | 4.2 | 1095 | 14 US-10-012-896-780 | Sequence 780, App |
| 37 | 110.5 | 4.2 | 425 | 9 US-09-815-242-10467 | Sequence 10467, A |
| 38 | 110.5 | 4.2 | 425 | 12 US-10-287-274-428 | Sequence 428, App |
| 39 | 110.5 | 4.2 | 476 | 10 US-09-738-626-5741 | Sequence 5741, Ap |
| 40 | 110 | 4.2 | 927 | 12 US-10-032-585-7923 | Sequence 7923, Ap |
| 41 | 109.5 | 4.2 | 665 | 10 US-09-942-447-2 | Sequence 2, Appl1 |
| 42 | 108.5 | 4.1 | 511 | 15 US-10-156-761-11421 | Sequence 11421, A |
| 43 | 108.5 | 4.1 | 581 | 9 US-09-815-242-5694 | Sequence 5694, Ap |
| 44 | 108.5 | 4.1 | 596 | 9 US-09-815-242-5244 | Sequence 5244, Ap |
| 45 | 108.5 | 4.1 | 603 | 9 US-09-815-242-12667 | Sequence 12667, A |
| 46 | 108.5 | 4.1 | 320 | 9 US-09-815-242-12525 | Sequence 12525, A |
| 47 | 108 | 4.1 | 604 | 15 US-10-156-761-92579 | Sequence 9297, Ap |
| 48 | 108 | 4.1 | 451 | 9 US-09-815-242-5079 | Sequence 5079, Ap |
| 49 | 108 | 4.1 | 933 | 9 US-09-815-242-11817 | Sequence 11817, A |
| 50 | 107 | 4.1 | 527 | 10 US-09-881-752A-136 | Sequence 136, App |
| 51 | 107 | 4.1 | 1043 | 10 US-09-738-626-6647 | Sequence 6647, Ap |
| 52 | 106.5 | 4.0 | 475 | 10 US-09-738-626-6236 | Sequence 6236, Ap |
| 53 | 106 | 4.0 | 530 | 15 US-10-156-761-13471 | Sequence 13471, A |
| 54 | 105.5 | 4.0 | 360 | 12 US-09-826-509-443 | Sequence 443, App |
| 55 | 105.5 | 4.0 | 378 | 12 US-09-826-509-445 | Sequence 445, App |
| 56 | 105.5 | 4.0 | 380 | 12 US-09-826-509-441 | Sequence 441, App |
| 57 | 105.5 | 4.0 | 387 | 12 US-10-118-661-23 | Sequence 23, Appl1 |
| 58 | 105.5 | 4.0 | 387 | 12 US-09-826-509-436 | Sequence 436, App |
| 59 | 105.5 | 4.0 | 388 | 9 US-09-989-861-8 | Sequence 8, Appl1 |
| 60 | 105.5 | 4.0 | 388 | 12 US-09-826-509-439 | Sequence 439, App |
| 61 | 105.5 | 4.0 | 388 | 15 US-10-157-031-123 | Sequence 123, App |
| 62 | 105.5 | 4.0 | 388 | 15 US-10-157-031-124 | Sequence 124, App |
| 63 | 105.5 | 4.0 | 388 | 15 US-10-225-567A-18 | Sequence 18, Appl1 |
| 64 | 105.5 | 4.0 | 489 | 15 US-10-156-761-11766 | Sequence 11766, A |
| 65 | 105.5 | 4.0 | 502 | 15 US-10-156-761-8917 | Sequence 8917, Ap |
| 66 | 105.5 | 4.0 | 531 | 12 US-10-154-419-81 | Sequence 81, Appl1 |
| 67 | 105.5 | 4.0 | 531 | 12 US-10-146-733-31 | Sequence 31, Appl1 |
| 68 | 105 | 4.0 | 427 | 10 US-09-922-501-6 | Sequence 6, Appl1 |
| 69 | 105 | 4.0 | 451 | 10 US-09-738-626-3815 | Sequence 3815, Ap |
| 70 | 105 | 4.0 | 490 | 9 US-09-739-457-3 | Sequence 3, Appl1 |
| 71 | 105 | 4.0 | 490 | 9 US-09-739-457-6 | Sequence 6, Appl1 |
| 72 | 105 | 4.0 | 490 | 9 US-09-739-457-7 | Sequence 7, Appl1 |
| 73 | 105 | 4.0 | 490 | 12 US-09-849-138-28 | Sequence 28, Appl1 |
| 74 | 105 | 4.0 | 490 | 12 US-09-849-138-43 | Sequence 43, Appl1 |
| 75 | 105 | 4.0 | 643 | 16 US-10-080-170-626 | Sequence 626, App |
| 76 | 105 | 4.0 | 1104 | 12 US-10-171-319-8 | Sequence 8, Appl1 |
| 77 | 105 | 4.0 | 2261 | 11 US-09-984-827-132 | Sequence 132, App |
| 78 | 104.5 | 4.0 | 786 | 9 US-09-815-242-10103 | Sequence 10103, A |
| 79 | 104 | 4.0 | 479 | 12 US-10-338-075-937 | Sequence 937, App |
| 80 | 103.5 | 3.9 | 572 | 12 US-10-154-419-13 | Sequence 13, Appl1 |
| 81 | 103.5 | 3.9 | 572 | 10 US-09-919-781-2 | Sequence 2, Appl1 |
| 82 | 103.5 | 3.9 | 772 | 10 US-09-738-626-6661 | Sequence 6661, Ap |
| 83 | 103 | 3.9 | 410 | 9 US-09-767-041-20 | Sequence 20, Appl1 |
| 84 | 103 | 3.9 | 438 | 9 US-09-815-242-5203 | Sequence 5203, Ap |
| 85 | 103 | 3.9 | 469 | 9 US-09-925-302-627 | Sequence 627, App |
| 86 | 103 | 3.9 | 537 | 15 US-10-156-761-12236 | Sequence 12236, A |
| 87 | 103 | 3.9 | 550 | 15 US-10-272-613-8 | Sequence 8, Appl1 |
| 88 | 103 | 3.9 | 2261 | 11 US-09-984-827-128 | Sequence 128, App |

```
Sequence 14001, A
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 8, Appli
Sequence 9, Appli
Sequence 4, Appli
Sequence 8, Appli
Sequence 2, Appli
Sequence 59, Appli
Sequence 14861, A
Sequence 2, Appli
```

DNA fragments and uses thereof

| Seq | As | Gaps | 0 |
|-----|-----------------|------|---|
| 1 | TTCTKTLTIVAGP | 60 | |
| 2 | TTCTKTLTIVAGP | 60 | |
| 3 | TTCTKTLTIVAGP | 60 | |
| 4 | TTCTKTLTIVAGP | 60 | |
| 5 | TTCTKTLTIVAGP | 60 | |
| 6 | TTCTKTLTIVAGP | 60 | |
| 7 | TTCTKTLTIVAGP | 60 | |
| 8 | TTCTKTLTIVAGP | 60 | |
| 9 | TTCTKTLTIVAGP | 60 | |
| 10 | TTCTKTLTIVAGP | 60 | |
| 11 | TTCTKTLTIVAGP | 60 | |
| 12 | TTCTKTLTIVAGP | 60 | |
| 13 | TTCTKTLTIVAGP | 60 | |
| 14 | TTCTKTLTIVAGP | 60 | |
| 15 | TTCTKTLTIVAGP | 60 | |
| 16 | TTCTKTLTIVAGP | 60 | |
| 17 | TTCTKTLTIVAGP | 60 | |
| 18 | TTCTKTLTIVAGP | 60 | |
| 19 | TTCTKTLTIVAGP | 60 | |
| 20 | TTCTKTLTIVAGP | 60 | |
| 21 | TTCTKTLTIVAGP | 60 | |
| 22 | TTCTKTLTIVAGP | 60 | |
| 23 | TTCTKTLTIVAGP | 60 | |
| 24 | TTCTKTLTIVAGP | 60 | |
| 25 | TTCTKTLTIVAGP | 60 | |
| 26 | TTCTKTLTIVAGP | 60 | |
| 27 | TTCTKTLTIVAGP | 60 | |
| 28 | TTCTKTLTIVAGP | 60 | |
| 29 | TTCTKTLTIVAGP | 60 | |
| 30 | TTCTKTLTIVAGP | 60 | |
| 31 | TTCTKTLTIVAGP | 60 | |
| 32 | TTCTKTLTIVAGP | 60 | |
| 33 | TTCTKTLTIVAGP | 60 | |
| 34 | TTCTKTLTIVAGP | 60 | |
| 35 | TTCTKTLTIVAGP | 60 | |
| 36 | TTCTKTLTIVAGP | 60 | |
| 37 | TTCTKTLTIVAGP | 60 | |
| 38 | TTCTKTLTIVAGP | 60 | |
| 39 | TTCTKTLTIVAGP | 60 | |
| 40 | TTCTKTLTIVAGP | 60 | |
| 41 | TTCTKTLTIVAGP | 60 | |
| 42 | TTCTKTLTIVAGP | 60 | |
| 43 | TTCTKTLTIVAGP | 60 | |
| 44 | TTCTKTLTIVAGP | 60 | |
| 45 | TTCTKTLTIVAGP | 60 | |
| 46 | TTCTKTLTIVAGP | 60 | |
| 47 | TTCTKTLTIVAGP | 60 | |
| 48 | TTCTKTLTIVAGP | 60 | |
| 49 | TTCTKTLTIVAGP | 60 | |
| 50 | TTCTKTLTIVAGP | 60 | |
| 51 | TTCTKTLTIVAGP | 60 | |
| 52 | TTCTKTLTIVAGP | 60 | |
| 53 | TTCTKTLTIVAGP | 60 | |
| 54 | TTCTKTLTIVAGP | 60 | |
| 55 | TTCTKTLTIVAGP | 60 | |
| 56 | TTCTKTLTIVAGP | 60 | |
| 57 | TTCTKTLTIVAGP | 60 | |
| 58 | TTCTKTLTIVAGP | 60 | |
| 59 | TTCTKTLTIVAGP | 60 | |
| 60 | TTCTKTLTIVAGP | 60 | |
| 61 | TTCTKTLTIVAGP | 60 | |
| 62 | TTCTKTLTIVAGP | 60 | |
| 63 | TTCTKTLTIVAGP | 60 | |
| 64 | TTCTKTLTIVAGP | 60 | |
| 65 | TTCTKTLTIVAGP | 60 | |
| 66 | TTCTKTLTIVAGP | 60 | |
| 67 | TTCTKTLTIVAGP | 60 | |
| 68 | TTCTKTLTIVAGP | 60 | |
| 69 | TTCTKTLTIVAGP | 60 | |
| 70 | TTCTKTLTIVAGP | 60 | |
| 71 | TTCTKTLTIVAGP | 60 | |
| 72 | TTCTKTLTIVAGP | 60 | |
| 73 | TTCTKTLTIVAGP | 60 | |
| 74 | TTCTKTLTIVAGP | 60 | |
| 75 | TTCTKTLTIVAGP | 60 | |
| 76 | TTCTKTLTIVAGP | 60 | |
| 77 | TTCTKTLTIVAGP | 60 | |
| 78 | TTCTKTLTIVAGP | 60 | |
| 79 | TTCTKTLTIVAGP | 60 | |
| 80 | TTCTKTLTIVAGP | 60 | |
| 81 | TTCTKTLTIVAGP | 60 | |
| 82 | TTCTKTLTIVAGP | 60 | |
| 83 | TTCTKTLTIVAGP | 60 | |
| 84 | TTCTKTLTIVAGP | 60 | |
| 85 | TTCTKTLTIVAGP | 60 | |
| 86 | TTCTKTLTIVAGP | 60 | |
| 87 | TTCTKTLTIVAGP | 60 | |
| 88 | TTCTKTLTIVAGP | 60 | |
| 89 | TTCTKTLTIVAGP | 60 | |
| 90 | TTCTKTLTIVAGP | 60 | |
| 91 | TTCTKTLTIVAGP | 60 | |
| 92 | TTCTKTLTIVAGP | 60 | |
| 93 | TTCTKTLTIVAGP | 60 | |
| 94 | TTCTKTLTIVAGP | 60 | |
| 95 | TTCTKTLTIVAGP | 60 | |
| 96 | TTCTKTLTIVAGP | 60 | |
| 97 | TTCTKTLTIVAGP</ | | |

421 EMATIPIDQOKYKGAIDVVAARPKSGALLGQGLVTCSSIGAMTYLAVILFII 480
421 EMATIPIDQOKYKGAIDVVAARPKSGALLGQGLVTCSSIGAMTYLAVILFII 480
481 AIWIVSATKINKLFLAOSALKEQVAQEDSAPASS 515
481 AIWIVSATKINKLFLAOSALKEQVAQEDSAPASS 515

e 3617 Application US/09738626
 tion No. US20020197605A1
 INFORMATION:
 INVENTOR: NAKAGAWA, SATOSHI
 INVENTOR: MIZOGUCHI, HIROSHI
 INVENTOR: ANDO, SEIKO
 INVENTOR: HAYASHI, MIKIRO
 INVENTOR: OCHIAI, KEIKO
 INVENTOR: YOKOI, HARUHIKO
 INVENTOR: TATEISHI, MAKIO
 INVENTOR: SENOH, AKIHITO
 INVENTOR: IKEDA, MASARO
 INVENTOR: OKAZKI, AKIO
 INVENTOR: NOVEL POLYNUCLEOTIDE
 REFERENCE: 249-125
 T APPLICATION NUMBER: US/09/738,626
 T FILING DATE: 2000-12-18
 APPLICATION NUMBER: JP 99/377484
 FILING DATE: 1999-12-16
 APPLICATION NUMBER: JP 00/159162
 FILING DATE: 2000-04-07
 APPLICATION NUMBER: JP 00/280968
 FILING DATE: 2000-08-03
 OF SEQ ID NOS: 7059
 RE: PatentIn ver. 3.0
 NO 3617
 H: 538
 PRT
 ISN: Corynebacterium glutamicum
 626-3617

| | | | | | | | |
|-------------------|--|----------|--------------|--------|------|--------|------|
| Accession | 6.28; | Score | 162; | DB | 10; | Length | 538; |
| Cal Similarity | 21.14; | Prod. | No. 1.1e-06; | | | | |
| 127; Conservative | 83; | Mismatch | 210; | Indels | 182; | Gaps | 29; |
| 2 | TKTEEK-----PIGKLR-----SELPWIRH--HELKVKLMEFLMFCIGTENTYVLRDXTD | 50 | | | | | |
| 6 | SKISERQGPAPFISKLSKMSDKRLNGVELTGNKLPPPLFILFILFITLALASSIMAMNV | 65 | | | | | |
| 51 | LIYAGPS-----GAEALPFIKFVLVPCAL-----IFM | 79 | | | | | |
| 66 | SVI--VPSSDELFVKGFLTEGGLTWLITNLGNAYIGFPLLTVLPLILAVGAERSGMIA | 124 | | | | | |
| 80 | LIYAKLSNLSKOALFFYAVG---TPFLIFPALFPTVIRPLRDV-----HPTFADRL | 129 | | | | | |
| 125 | ALIKKLGSKKIVLPLPAVGVIGTASIMADAAF--VVEPPLAAMPKKAAGRHPV----- | 177 | | | | | |
| 130 | QALIPGLGLVALIRNMTFAAFYVLAELWGSVMLSLMFGEFANEI-----TKIHE | 180 | | | | | |
| 178 | -----AGLTSPFAV-----GAGYSTAIV--PTSLDALPAGITNAVMTPLPGIATTEVNP | 225 | | | | | |
| 181 | AKRFPALFEGIGANISL--LASGRAI-----IWASKLRASVSEGVDPWG-----ISURL | 226 | | | | | |
| 226 | VSNY--FNIASSIVGLGCGFLIDKVLBPRMRQKIATEVASIEEPLISAADDEISATL | 283 | | | | | |
| 227 | -----LMAMTVSGVLVMAISYMWINKNVITDPRFYNPDEMQKKGAKPKR | 271 | | | | | |
| 284 | TAGENRALTISMWTLTATAIIVLVVLILGSPRRNED-----GGFLPT | 326 | | | | | |
| 272 | NNMKDSGLVYDRSPYILLTLVLIVANGICNLIEVTWKSQKLQYIPNNNDVSEFPNGNSF | 331 | | | | | |
| 327 | SPLLSSVVF-----VFLEFVWMLAAGNVVGITK-----NDDDVNNMGAL | 369 | | | | | |

```
Qy 332 WTGVSVLIMEVGGNVIRKFGMLTGLALVTPVMTLLTGIVFPALVIFRNOASGU--VAMF 389
Db 370 -KDMIGFLVLAIFILGQFALFRW-----TIGTWAV--QGAIGLEAIGIT 412
Qy 390 GTPEMLAVVGAION--ILSKSTKYALFDSTKEMAYIPLDOEKVKKAAIDVVAARFG 447
Db 413 GFPALIAFIIILASCLNIIISGSAMWTL-----MAAVFVPMFALLGYEPPFIQAARVAG 466
Qy 448 KSGGALI---OQGLLVTCGSGAMTP-----YLAVILLFIATLVATKLNKLFIAQS 498
Db 467 DSATQVITPLNPYMIIVILGLRRYEPDAGLTMSRLIPVIPEMLATMTLAIWPFYADL 526
Qy 499 AL 500
Db 527 PL 528

RESULT 3
US-09-815-242-10144
; Sequence 10144, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zykkind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10144
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10144

Query Match 5.0%; Score 132.5; DB 9; Length 463;
Best Local Similarity 22.3%; Pred. No. 0.00058;
Matches 108; Conservative 69; Mismatches 174; Indels 133; Gaps 28;

Qy 69 WLVPACAIIFMLIYAKLSNI--LSKQALFYAVGTPELIFALFPYIYPLRVLHPTERFAD 127
Db 39 WISVPCILLAFQVWMLPSAVAVNLPRKVGFNFTTQDLFMLTLPLPSVSGALLRV--PYSP-- 94
Qy 128 RLQALIPPLGLGLVAIL--RNMTPAFYVLAELMGSVWLSLMPFGPANEIRKIHAKRFY 185
Db 95 -----MPIFGGRMT--AFST-----GILIIPCVWIGFA-----VQDTSTPY 130
Qy 186 ALFISGANISLLASGRALVWASKLRASVS-----EGVDPMGISLRLLMAM 230
Db 131 SVFTI---ISLLCGFAGANFASMS-ANISFFPKQKQCGALGLNGCIGNMGVSVWQLVAP 186
```

```
Qy 231 TIISGLVIMASVWINKVLTDPFRFYNPEEMQKKAKPRGMNKKDSFLYDRSPYIL- 289
Db 187 LVVS-LSTFAVF-----DSQKQPDGTE--LYLANISWIVP 221
Qy 290 -LTLVIAVAGICINLIEVYTKSOLKQYNNND-----YSEFMGNF-SFWTGV--- 335
Db 222 FLAFTIAWFGNMDL-ATSKASIKEDQLPYLKRHLMIMSLYLATSGSFIGSAGFAML 280
Qy 336 -----VSULIMLFGV--GNVIRKFGW-----LTGALVTPV---MVLITGIVFPALV 376
Db 281 SKTQFPDVQILQYAFPGFPGIALARSGALSDRLIGTRVTLVNFIIIMAFSGILFLTLP 340
Qy 377 IFRNOASGLVAMFGTPEMLAVV-----GAIQNIISKSTKYALFDSTKEMAYIPLDOE 430
Db 341 T-DGQGSFMAFPA--VFLLALFLTAGLSGSTQMTSVIFRKLTMDEKAEQ---GSD 392
Qy 431 QKVGAKAIDVVA-----RFGKSGALLQOGL--LVTCGS-IGAMTPYLAIVILFIIA 481
Db 393 ERARREATDPAALGFIISALGALGFFIPKAFSSALITGSPVGAMKVFILFIYACVI 452
Qy 482 IWLV 485
Db 453 TWAV 456

RESULT 4
US-09-815-242-10161
; Sequence 10161, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zykkind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10161
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10161

Query Match 4.9%; Score 129; DB 9; Length 462;
Best Local Similarity 22.3%; Pred. No. 0.0012;
Matches 108; Conservative 71; Mismatches 171; Indels 134; Gaps 26;

Qy 69 WLVPACAIIFMLIYAKLSNI--LSKQALFYAVGTPELIFALFPYIYPLRVLHPTERFAD 127
Db 37 WISVPCILLAFQVWMLPSAVAVNLPRKVGFNFTTQDLFL--TLPLPSVSGALLRVYSPMV 94
```


Db 209 HHRFQGLSLCTCYASISIGGIATLT-----GTTFNVLQGGVNSLFPQNGVNVFASMF 262
Qy 327 GNFSGWTVSVLT-----MLFVGNVIRKRGW-----LTGAL- 359
Db 263 G-FAFPTMILLLLAMLMQLVFLGVNFRKNFGFGEGERKQAAFOVKTQYRLIGMS 321
Qy 360 -VFVWVLLTGTIVF-----FALVIFRQASGLVAMFGTTEPLMAVVGAION 405
Db 322 FAEKTVTLFVLVLTFTREBGFPPGMDIVFANEKQSMASDGTVAIFISLVFTIP- 380
Qy 406 ILSKSTKXALFDTKEMAYIP-LDQOKVYKKA-----AIDVVAR 445
Db 381 -----SKTIGLMQDPKKPKLKAAPALITWKTVNDKMPNIVILLGGG 423
Qy 446 FKGSGGALIQGLVTCISIGAMTPY-----LAVILFTIATLV-----VSATKLNKL 493
Db 424 FALAKGS-EQSGLSMLGD-KLTPLOHIPESATVILCLIAIFTECTSNVATTLFLP 480
Qy 494 FLAOSA 499
Db 481 ILASMA 486

RESULT 7
US-10-014-717-11
; Sequence 11, Application US/10014717
; Publication No. US2002019278A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zitzke, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-10-014-717-11

Query Match 4.7%; Score 122.5; DB 14; Length 713;
Best Local Similarity 21.2%; Pred. No. 0.0091;
Matches 105; Conservative 67; Mismatches 165; Indels 159; Gaps 21;
Qy 76 IIFMLIYAKLNLNLSKQ-----ALFYAVGTPLIFALFPYIVPLRDVLTHTPEPAD 127
Db 16 LALLIVTRARAGSELARRRQPEVIGELRGV-----VLGFSVGLAAGCFHRAFOE 67
Qy 128 RLQAILPGLIGLVAILRNMTFAFYVLAELMGVSLMFWGANETIKIHEARKFYAL 187
Db 68 PAVGVVLSGI-----SWIGALLILLMAGIEVDVIGIRKEARFAL 107
Qy 188 FGIGANISLASGRAIVWASKLRASVSEGVDPWGISLTLAMMTVLSLV-----LM 239
Db 108 SALGAIAPPLAAGA-----FSALVLDRLPSPSGLFIIVSVTAIVSVIANVLIERESMR 161
Qy 240 ASYVWIN--KNVLEDPFRFYNPEEMQGGKAKPKXNMDSFLYLDPSYIILLTLVAY 297
Db 162 RSYAQVTLAAGVSVAAWLVAMTSSSYGASPALAVARSALL--ASGFLFPMVLV----- 215
Qy 298 GICINLIEVTW-----KSQALKIQYNNNDYSEKMGNSFWTGVSVLIMLFGVGNVI 349
Db 216 GRRLTLHLMRVVADATRVSKQGVSL-----VLVLTFLAALTL 252

Qy 350 RKFGW--LTGALVTPVMV-----LTGTI-----VEFALVIFRQASGL--V 386
Db 253 QRLGHPILGAPFALGVNLNSAPRTNRPILDSVQLVAGLPAPVFFVLAGMVDVSQLETP 312
Qy 387 AMFGTTPMLA-----VVGAIONILSKSTKVALFDTKEMAYIPLDQOKVYKKAID 440
Db 313 AAMGTVALLLATATVAAKVPALGARLG-----GLRGEALVAVGLNM-----KGGTD 361
Qy 441 VVAARFGSGGALIQGLVTCISIGAMTPYLAIVLTFTIA-----IMVSRATKLNKLFL 495
Db 362 LIVAIVG-----VELGIL-----SNAVTVYAVVALVTASPALLLWLE----- 401
Qy 496 AQSALKEQEVADQSDA 511
Db 402 -----KRAPPTQESHA 412

RESULT 8
US-09-815-242-13892
; Sequence 13892, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zvekind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Xu, H. Howard
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: EUTRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: PasteSeq for Windows Version 4.0
; SEQ ID NO 13892
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13892

Query Match 4.6%; Score 120.5; DB 9; Length 395;
Best Local Similarity 20.7%; Pred. No. 0.0064;
Matches 100; Conservative 71; Mismatches 196; Indels 117; Gaps 20;
Qy 28 LPMFLMFCITENTYVLRDRTDILIVGAPSGAAIPIFKWLVY--PCAIIFMLIYAKL 85
Db 3 LCPMLALFALLISFAI-GTTEFVIQV-----LVPTIAQQLAVSIPSAGMLVLIY-- 51
Qy 86 SNLSKOALFYAVGTPLIFALFPYIVPLRDVLTHTPEF--ADRLOATILPGLIGIVA 142
Db 52 -----ALGVATGAPVLT--AL--TGRULPRKQULVALMLFTKGNLLAMQAPGYMTLLIA 100
Qy 143 ILRNWTPAIFYVLAELMGVSLMFWGANETIKIHEARKFYALFGIGANISILAS-- 199
Db 101 -ARLLTGLAGVFPFISISITATSL-----VPEKKAASAIAMFG-GILVALVTGVPL 150

QY 200 -----GRAIWMASKLRASVSEGVDPWGISLRLLMAMTIVSGVLIMASYMMINKNVLTPDRF 255
 DB 151 GTTICQHGWR-----ETFLAVSLGLVILMSSQLIPANT----- 186
 QY 256 YNPEEMQKKGKAKPKNMKDSFLYLDSPYLLLTLLVAVGICINILEVTWKSQKLQ 315
 DB 187 -----GPRAAA-----SIRDOVKVLTFRLLIYAVALAGYGVF-----TAFTEL 227
 QY 316 YPNNDVSEFMKNFSEFWGV---VSUMLMFLVGVGVIRKFGMLTGALTPVAVLLTGIVF 372
 DB 228 APMMODLGFSPAVASWILLGVSVAILGNWIGKLDKXG-----AVPALKTFPALF 281
 QY 373 FALVIFRNQASGLVAFGTTPLMLAVVGAIONISKSTKVALPDKSTEMAYIPLDORQ- 431
 DB 282 VLMWFQYLTASQYALATILVMGIFAGCNVPGI-----QYVVOGKAKQF 326
 QY 432 ---KYKRAIDVVARFGKSGGALIQQGLVTCSSIGAMPYLAIVLLPITATLVSAT 488
 DB 327 TPNADVAVASGNIAAFNIGIALGVIGQTVAHYGL--AQTFWIGALIVLAVFLIMGVSG 384
 QY 489 KLNK 492
 DB 385 RLDK 388

RESULT 9
 US-09-815-242-12171
 ; Sequence 12171, Application US/09815242
 ; Patent No. US2002061569A1
 ; GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyekind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 12171
 LENGTH: 425
 TYPE: PRT
 ORGANISM: *Staphylococcus aureus*
 US-09-815-242-12171

Query Match 4.6%; Score 120.5; DB 9; Length 425;
 Best Local Similarity 19.5%; Pred. No. 0.0071;
 Matches 94; Conservative 86; Mismatches 160; Indels 143; Gaps 19;
 QY 9 FGKRSFIMPIHTHELEKLVLPMEFLMFCTITENYTVLTKDTLLIVAGDAGAGAI----- 63
 DB 7 FGAGNLIPPPMLGHTAGCNMMWIGMLGFALT---GILLPFIIVIVAFYDEGVESVGNRIH 63

QY 64 PFIFKWLVPICATIFMLIYA-----KLSNLSKQALFYAVGT-----PFLIFPAL 108
 DB 64 PMPGFIKAV---VYMSIGAFYGIIPRAANVA-----YEICTRHLPVHNQWTLIIIFAAV 114
 QY 109 FPTYIPLRDVLAHTEPADRLOAILPGLGLVAILRNWTPAAY----- 153
 DB 115 FFAIVYWS--LNSPKYVDNIGKLLTPLLLMVALL---SIAYFNBPASALAPKDKYT 169
 QY 154 -----VLAELMGSYML-----SLMFWGANEITKIHAKRF 184
 DB 170 HPIFSGSIEGYFTMDVLAALFVVIYNGYFKGLTDRTKLKYVCBSGFIAILLIGMIV 229
 QY 185 YALFGIGNIS--LLASGRAIWMASKLRASVSEGVDPWGISLRLLMAMTIVSGVLIMASY 242
 DB 230 FALAVYGASTPAGNFKQGTIDILTVNSLRVFGSGFNLVFGMTV--ILACITCIGIV-NACA 287
 QY 243 WINKNVLTPDRFNPPEEMQKKGAKPKNMK-----DSFLYLDSPYILL--LT 291
 DB 288 TFKKHV-----PKFSYKIFALVFSIIGFLFTTLGLEMLIKIAVP 327
 QY 292 LLYVAGICINILEVTWKSQKLQYPMNDVSEFMGNFSPWTGVSVLIMLFVGNAVIRK 351
 DB 328 LTLIYFVSIALVLSF-----ANMFSTRFSAVRLATVITLIIISIL--QILNS 375
 QY 352 FGMULTGALVTPVAVLITGIVFALVIFRNQASGLVAFGTTPLMLAVVGAIONISKST 411
 DB 376 FNLHGVLKMKFMLPLADIDLAVL-----PFLILAIIGTIDIVAIRP 420
 QY 412 KYA 414
 DB 421 KQA 423

RESULT 10
 US-09-815-242-13831
 ; Sequence 13831, Application US/09815242
 ; Patent No. US2002061569A1
 ; GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyekind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 13831
 LENGTH: 466
 TYPE: PRT
 ORGANISM: *Salmonella typhi*
 US-09-815-242-13831

Query Match 4.5%; Score 118; DB 9; Length 466;
 Best Local Similarity 20.4%; Pred. No. 0.014;
 Matches 99; Conservative 70; Mismatches 179; Indels 138; Gaps 23;

69 WLVPVCAIIFMLIYAKLSNI---LSKQALFYAVGTPPLIFALPFTVIYPLRDVLHPTTF 125
 41 WISVSCILLARCVMMFLSAVAVNLNKIGFDTTDLFIL-----TAL 82
 126 ADRLOATIPRGLGLVAL--RNTTFAAFVYLAELMGVMSLMWGA--NEITKHEA 181
 83 PSLSGAILRVSEFVPLFGGRKMTVLSVIL-----IIPCAMLGFAVGNPATP----- 131
 182 KRFPALFGIGANISLASGRAIWMASKL-----RASVSEGVDPWGISRLT 227
 132 -----FQVFLILALTCFAGANFASMGNISFPFPRKQSGALCINGLNGVSWQL 185
 228 MAMTIV-----SGLVLMASVW-WIN-KNVLDPDPFYNPEMOKGKKG 267
 186 IAPVIFPIFTPLGVRGVPQPDGSLALVTNAAMITVPLAVATLAAMFGMNDIGSSKAS 245
 268 AKPKNMDSFLYIDRSYIILLTLVIA-YGICINLIEVTWKSQKIQYRNM-----D 321
 246 VA-----SQLPVKRLHMLSLYLATGSGFIG-ESAGFAMLAQTQPDVNILOAF 297
 322 YSEFMGNFSFWTGVSVLIMLFVGNAVIRKFGMLTGALVTPV-WVLITGIVFPAIVFRN 380
 298 FGPIG-----ALASAGVIVDSKFGGVAVTLINIFMALPFLALFPLT---PG 343
 381 QASGLVAMF-----GTPPLMAVVGAIQNLISKSTKY--ALPSTKEMA 423
 344 SGASFSFAFLVFMGLFTAGLGSGSTQMTAVIPRO-----TLVNVLRGSDDEQA 396
 424 YIPDQEKVKGAIDVVAARFGSGGALLQGL---LVICGS-IGMPTYLAVIILFT 479
 397 -----QRAVADTDAALGISAIGAVGGFIPKAGTSLATGSPVGMKIFLLEYIACV 451

QY 480 IAIMLV 485
 Db 452 LITMLV 457

RESULT 11
 US-09-815-242-5678
 ; Sequence 5678, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zykend, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR APPLICATION NUMBER: 2001-03-21
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5678
 ; LENGTH: 430
 ; TYPE: PRF
 ; ORGANISM: Staphylococcus aureus
 US-09-815-242-5678

Query Match 4.5%; Score 117.5; DB 9; Length 430;
 Best Local Similarity 19.3%; Pred. No. 0.014;
 Matches 93; Conservative 86; Mismatches 159; Indels 143; Gaps 19;

9 FGKRSFLMPPIHTELEKVLPMFLMPCITFENVTLBDTKOTLIVGARGSAEAI----- 63
 15 FGAANLIFPMLGHTAQQNMWIGLGFALT---GILPFTIYIYAVAYDEVESEVGRKH 71
 72 PMFGFIPAV---VIYMSIGAIFYGIPRAANYA-----YEIGTRHILPVHNMOTLILFAV 122
 64 PFIFMLVPCAIIFMLIYA-----KLSNLSKQALFYAVGT-----PFLIFPAL 108
 109 PFTYIYPLRDVLHPTERADRLOALPRLGLGLVALILNMWTFAPY----- 153
 123 FFAIVYMS--LNBSKIVDMGLTLPLLLMALL---SIAYIFNESALSAPKDKYIT 177
 154 -----VLAELMGVYML-----SLMFWGANETIKIHEAKRF 184
 178 HPTISGSLGFTMDVLAALAFSVIYVNGYKFKLTTRTKILKYVCSGFIAILLGMY 237
 185 YALFGIGANIS--LLASGRAIWMASKLASVSEGVDPWGISRLMAMTIVSGVLMASY 242
 238 FALAYVGSAPGNFKDGTDLITVNSLRVFGSGNLVFGMTV-ILACLTTCIGLV-NACA 295
 243 WVIKNVLTDRPFYNPEMOKGKGAAPKXNMK-----DSLVIYDRSYIILL--LT 291
 296 TETKGVH-----PKTSYKIFALVFSIIGFLFTTLGHEMLIKIAPV 335
 292 LTVIAYGICINLIEVTWKSQKIQYRNMNDYSEFMGNFSFWTGVSVLIMLFVGNAVIRK 351
 336 LITLIVPSIALVLISF-----ANMFSTFRFSAKRLATVTLIISIL--QILNS 383
 352 FGMLTGALVTPVWVLITGIVFPAIVFRNOASGLVAMFGTTPMLMAVVGAIQNLISKST 411
 384 FNLHGIVLKMFWMLPLADIDLAMLV-----PFLLMALIGTILDIVVIRP 428

QY 412 K 412
 Db 429 K 429

RESULT 12
 US-09-815-242-13840
 ; Sequence 13840, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zykend, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR APPLICATION NUMBER: 2001-03-21
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26

```

; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13840
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(477)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13840

```

Query Match 4.4%; Score 116.5; DB 9; Length 477;

Best Local Similarity 20.2%; Pred. No. 0.02; Matches 95; Conservative 84; Mismatches 186; Indels 105; Gaps 27;

```

QY 69 WLVPICATIFMLIVAKLSNI-LSKQALFYAVGTPFLIFPALPTVIYPLRDVLPTEPAD 127
DB 51 MISVCCILAPCVMLFSAVANLPRKGFNTTDDLMFLTALPSVSGALLVPPY----- 104
QY 128 RLQALIPGLGLVALINMTFAFVYLAELGWSVLSLMEWGFANEITKHEAKRYAL 187
DB 105 -SKWVP--LFG---GRRWT--AFST-----GILIVPCWLGXA-----VQDTSTPST 144
QY 188 FGIGANISILLASGRAIWAASKIARASVS-----EGVDPMGISRLLMATTI 232
DB 145 FII---ISLLCKXXGAXPASM-ANISFFPRKQGGALINGLGNNGVSVMLXAPLV 200
QY 223 V-----SGVLMAASYW-WIN-KXVLTDPFVYPEEMQKKGAKPRM 272
DB 201 VLSLTFAPAGSHGVEQPDGSQLYLANAMWIPFLAITTLAIFGMEL-----ATSKA 254
QY 273 NMKDSFLYIDRSPLYLLTLTLVIA-YGICINLIEVTWKSQKLQYPMNDYSEFMGNFSF 331
DB 255 SLKEQLPVLRKG-HIMIMSLVLTATFGSFIG-FSAGFAMLSKTQPPDVQ-----ILHAF 307
QY 332 WTGVSVVIMLFVGNVIRKRGMLTALVTPY-MULLIGIYFPALVIRNQAASGLVAMEG 390
DB 308 FGPFGLALAR-SAGGALISDRUGTGTVTLLINFLVLAIFSGLLFLTLP-GGVGGSEIAPFG 365
QY 391 TTPMLAVVV-----GAIQNLISKSTKYALFDSKEMAYIPLDDEQKVKGAADVVAA 444
DB 366 ---VFLAFLTRAGLGSGSTFQMIISVIFPKLTMDRKAG-----GSDDEAMREAAITDAAA 418
QY 445 ----RFGSGGALIQQGL---LVICGS-IGAMTPYLAIVILLFTIAVLV 485
DB 419 LGFISALGAIIGFFIPKAFGSSSLALTGSPVGMKFLIFYIACVITWAV 468

```

RESULT 13

```

US-09-738-626-3861
; Sequence 3861, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

```

```

; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18 377484
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3861
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3861

```

Query Match 4.4%; Score 115.5; DB 10; Length 410;

Best Local Similarity 22.8%; Pred. No. 0.02; Matches 72; Conservative 47; Mismatches 104; Indels 93; Gaps 16;

```

QY 157 ELKGSVWLSL-MFWGFANEITKI--HEAKRYA-----LFGIGANISLLAG-- 200
DB 20 KVMILAVLSVFTVAMG-GNEFTPLLVFVRGSGFNSLFDLLVFAIGYAVGLLAGPL 78
QY 201 -----RAIVASKLARSVSGVDPMGISRLLMAM-----TIVSGVLMAASYWINKV 249
DB 79 SDRYGRRAVWMLPALIILLSGALIASGEETAILAIQRVUSGISVGNWIRGSGWIRE-- 136
QY 250 LTDRFYNPPEEMQKKGAK-PRKNNKDSPL-----YLDSPYILLTLV 294
DB 137 LSSSRF---EFGVTSAGAKRASWSLTGFGALGALGVAQMLPLPGQALVYVHLITLL 193
QY 295 IAYGICINLIEVTWKSQKLQYPMNDYSEFMGNFSWTGV--SVIMLFVGNVIRKF 352
DB 194 ILFPILITABETROSALHKTK-----GSFWSVDLVPSALDKRFL--FVAVPI 238
QY 353 G-WLTGLVTPVWVL-----LTGIVFPALVIFRNQASG----- 384
DB 239 GPWFGLAFTAYAVLPSQLDMSAPAVYSHLIALVLTLSGFGIQFGPQIMGSKTRGP 298
QY 385 LVAMEGTTPLMLAVVV 400
DB 299 ILAMFTVIGMIGAVI 314

```

RESULT 14

```

US-10-181-319-4
; Sequence 4, Application US/10181319
; Publication No. US20030135032A1
; GENERAL INFORMATION:
; APPLICANT: Lewis, Thomas A.
; APPLICANT: Paszczynski, Andrzej
; APPLICANT: Crawford, Ronald L.
; APPLICANT: Cortese, Marc S.
; APPLICANT: Sebati, Jonathan L.
; TITLE OF INVENTION: Compositions and Methods for Bioremediation
; FILE REFERENCE: IDAHI19440
; CURRENT APPLICATION NUMBER: US/10/181,319
; CURRENT FILING DATE: 2002-07-15
; PRIOR APPLICATION NUMBER: PCT/US01/02386
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/177,251
; PRIOR FILING DATE: 2000-01-20
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Pseudomonas stutzeri
US-10-181-319-4

```

Query Match 4.4%; Score 115; DB 12; Length 392;

Best Local Similarity 20.9%; Pred. No. 0.021;
Matches 88; Conservative 55; Mismatches 127; Indels 152; Gaps 20;

```
QY 94 LEVATGPTPLIFALFPIYIVLRLVLTHTERPAD-----RLQ-ALPPLGLGVALIRMT 148
DB 18 LIIANGMPWMIYFAL-----GILGHLVADIGISROQLGWLTAFTGALILSPWA 68
QY 149 FA-----AFVYLAELMGSVMLSLMFPGFANEITKTHEAKRFYALFGIGANTSL 196
DB 69 GALVQRMGRACLICFLLVGLSFLMAVLP--GREGVLTAL-----LLCGTQOSLAN 119
QY 197 IASGRAIYASKL--RASVSEGVDPWGISLRLLMAMTVISGLVLM-----239
DB 120 PATNQALIASVVARAKAGVGLKQSGVQASALLAGALPVLVLMGMRGALAAVPAVALV 179
QY 240 ----ASYMINKVLTDPFNFPEMOKGKAKPRMNMKDSFLYDRSPYILLTLTLVY 295
DB 180 MAALVTYWPAAKSV-----SAPSLPLR-----VKGPNVWLSILMAI 215
QY 296 AVGICINIEVTWKSQKLQYPMNDYSBFMGNFSTWGV--VSVLJMLFVGNAVIRKF-- 352
DB 216 Q--LQAGLA-----LSSFMTFLGVYAAQIGSVSTI-----GAMVSCFA 253
QY 353 -GWLTCALVTPVM-----VLTGIVF-----PALVIFR--NQASGLVAMFGTTPMLAVV 399
DB 254 MGILSRVLLTPIADKLKOBETILLGVLFILAGLALVMRANQOHMPLMLGVTGMGLTV- 312
QY 400 VCAIWNILSKSTKYALPDSTKEMAYIPLDQEQVKGAALDIVAAPFGSGGALLIQGLL 459
DB 313 --AASNAIAMS-----MLRDGRFGGAATSAAGWL 339
QY 460 VI 461
DB 340 SV 341
```

RESULT 15

US-09-738-626-5024
; Sequence 5024, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.10
; SEQ ID NO 5024
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5024

Query Match 4.4%; Score 115; DB 10; Length 708;
Best Local Similarity 18.6%; Pred. No. 0.047;
Matches 106; Conservative 70; Mismatches 149; Indels 244; Gaps 25;

```
QY 109 FPTVYPLRDVLTHTERPADRIQALIPG-----LGLVALIRMTAFAYVLAEL 158
DB 120 YPHVRPAPAPAKPPTSEKIMRGVATIGGVITVAGVILLVSAIQRCH--LGPLGRV 175
QY 159 WGSVWLSLMFPGFANEITKTHEAKRFYALF-----GIGA 192
DB 176 ICAVYLAIVLIGAAHYVK--RGTFVEALVLTVSQIAFLATSAIIFILEMPPGIGS 233
QY 193 ----NISLASGAITYMA--SKLRASVSEGVDPWGISLRLLMAMTVISG-----LVYMA 240
DB 234 LVALLIGNIFLIVGR--LMSLSKTEKSAEG-----HTFVGAIAVSGFSAIIFALSA 284
QY 241 SYMINKVLTDPFNFPEMOKGKAKPRMNMKDSFLYDRSPYILLTLTLVAVGIC 300
DB 285 DAWM-----PISIVAAALLSIRIS 304
QY 301 INLI-----EYTWKSQ-----LKLQYPMNDY 322
DB 305 TIIIPASMAAFVILIQFLSASMQTMEWPATITGTTAVLLVALTMDPFXITATDSDHI 364
QY 323 S--EFMGNF-----SFMTGVSVLIMLFVGNVIRKFGMLTGAIVTPMVLITGVFAL 375
DB 365 ALBEYWRSPETNPVSTWGAIVSPVLIIVITTSMPFAVDMPMLALIPACAVVALGI--PAL 422
QY 376 ----VFRRQASGLVAMFG-----TTPMLAV--VGA-----402
DB 423 RSSDTASIEORMSRLVAVGLALIAETVQLFTGDLEFTNPLVWVPLIAGALLPMLRM 482
QY 403 ----IONILSKSTKY-----ALFDSKEMAYIPLDQEQ 431
DB 483 LPQRQLGVPEWVAMLIAAVMTGVLLENVVISIPMLMTDQALQALLIVFI-----536
QY 432 KVKGAALDIVAAPFGSGGALLIQGLLV--CGSIGAMPYL-----AVILFTI 480
DB 537 ----AATIQVRRSPFGHMLQIIVGLTLLTSAISIVITITFTIGRLIAGNAGMWLGRFI 592
QY 481 A-----IWLVSATKL--NKLFLAQSA 500
DB 593 CHATVSIIMWVIAAALMLNRLKLDLPAGL 621
```

RESULT 16

US-09-759-143-778
; Sequence 778, Application US/09759143
; Patent No. US20020022248A1
; GENERAL INFORMATION:
; APPLICANT: XU, Jiangchun
; APPLICANT: DILLON, David C.
; APPLICANT: MITCHEM, Jennifer L.
; APPLICANT: HARLOCKER, Susan L.
; APPLICANT: JIANG, Yugu
; APPLICANT: HENDERSON, Robert A.
; APPLICANT: KALOS, Michael D.
; APPLICANT: FANGER, Gary R.
; APPLICANT: RETTER, Marc W.
; APPLICANT: STOLK, John A.
; APPLICANT: DAY, Craig H.
; APPLICANT: VEDVICK, Thomas S.
; APPLICANT: CARTER, Darrick
; APPLICANT: LI, Samuel
; APPLICANT: WANG, Aijun
; APPLICANT: SKEIKY, Yasir A.W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759, 143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 778
; LENGTH: 1095
; TYPE: PRT

Qy 257 NPEEMQKGGKAKPRNMKDSFLYLDSPYLLTLTLVIAVGCINLI----- 304
Db 841 -----SRLNGKXIMLQRM- IDVFLEFLFAVMVAFGARQGLLRQNEGRWKIF 891
Qy 305 -EVTWKSQKL- -QYPMNDYSEFMGNFSFMTGVSVLIMLFGVGNVIRK- GMLTGALV 360
Db 892 RSVIYEPYAMFGVDPDVGDTYDFAHCTFTGNESKPLCVELDEHNLPRPEWITPLV 951
Qy 361 TPWVNLITGIVFPALVIFRNQASGLVAMFGTTPMLAVNGAIG 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVQ 978

RESULT 19

US-09-895-793-778
; Sequence 778, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hurai, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Basols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 778
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-778

Query Match 4.3%; Score 113; DB 10; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.13;
Matches 85; Conservative 54; Mismatches 121; Indels 144; Gaps 19;
Qy 43 VLBDTKDILLVAGAGSAGEAIPIKFWLVPCAIIFMLI-----YAKLSNLSKQALFY 96
Db 677 ISRDTKN-----WKIILCLFIPLVCGGFVSFRKKPVDKHKKLLMW 717
Qy 97 AVG---TPELI-----FPALFPTV-IYPLRDVILHPTFPADRLOALIPGLGLGIVA 142
Db 718 YVAFPTSPVAVFSNNVVFYIAFLFLFAYVLLMDPHSVHPPELV-----LysLVF 767
Qy 143 IL-----RNMTFAFYVLAELMGVMSL-MFWGFANEITRIHEAKRFYALFGIGANISL 196
Db 768 VLPCDEVRQWYVNGVNYFTDLN-NVMDTLGLFYFIAGIVRRLH-----SSKSS 815
Qy 197 LASGRAIWMASKLPAVSSEGVDPWGISLRLLMAMTYISGLVLMASVWINKNVLTDPRFY 256
Db 816 LYSGRVIFC-----LDYIIFTLRLIHIFTV----- 840

Qy 257 NPEEMQKGGKAKPRNMKDSFLYLDSPYLLTLTLVIAVGCINLI----- 304
Db 841 -----SRLNGKXIMLQRM- IDVFLEFLFAVMVAFGARQGLLRQNEGRWKIF 891
Qy 305 -EVTWKSQKL- -QYPMNDYSEFMGNFSFMTGVSVLIMLFGVGNVIRK- GMLTGALV 360
Db 892 RSVIYEPYAMFGVDPDVGDTYDFAHCTFTGNESKPLCVELDEHNLPRPEWITPLV 951
Qy 361 TPWVNLITGIVFPALVIFRNQASGLVAMFGTTPMLAVNGAIG 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVQ 978

RESULT 20

US-09-895-814-778
; Sequence 778, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hurai, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Basols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 778
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-814-778

Query Match 4.3%; Score 113; DB 10; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.13;
Matches 85; Conservative 54; Mismatches 121; Indels 144; Gaps 19;
Qy 43 VLBDTKDILLVAGAGSAGEAIPIKFWLVPCAIIFMLI-----YAKLSNLSKQALFY 96
Db 677 ISRDTKN-----WKIILCLFIPLVCGGFVSFRKKPVDKHKKLLMW 717
Qy 97 AVG---TPELI-----FPALFPTV-IYPLRDVILHPTFPADRLOALIPGLGLGIVA 142
Db 718 YVAFPTSPVAVFSNNVVFYIAFLFLFAYVLLMDPHSVHPPELV-----LysLVF 767
Qy 143 IL-----RNMTFAFYVLAELMGVMSL-MFWGFANEITRIHEAKRFYALFGIGANISL 196
Db 768 VLPCDEVRQWYVNGVNYFTDLN-NVMDTLGLFYFIAGIVRRLH-----SSKSS 815
Qy 197 LASGRAIWMASKLPAVSSEGVDPWGISLRLLMAMTYISGLVLMASVWINKNVLTDPRFY 256
Db 816 LYSGRVIFC-----LDYIIFTLRLIHIFTV----- 840

Qy 257 NPEMOKGKAGKAPKMMKDSFLYLDSPYLLTLTLVAVGICINLI----- 304
Db 841 -----SRNLGPKIIMLQRM-L-IDVFFFLFAVMMVAFGAROGILRQNEQRWIF 891
Qy 305 -EVTWKSQKL--QYPMNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKF-GMLTGALV 360
Db 892 RSVIYEPIYAMFGVPSVDVGTGYDPAHCTFTGNSKPLCVLDEHNLPRPEWITIPLV 951
Qy 361 TPVWVLLTGIVFALVIFRNQASGLVAMFGTTPMLAVVGAIQ 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVQ 978

RESULT 21

US-10-144-678A-778
; Sequence 778, Application US/10144678A
; Publication No. US20030157089A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, Daghun
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Rafter, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Hepler, William T.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals y de Bassols, Carlota
; APPLICANT: Foy, Teresa W.
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Deng, Ya
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C28
; CURRENT APPLICATION NUMBER: US/10/144,678A
; NUMBER OF SEQ ID NOS: 1033
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 778
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-678A-778

Query Match 4.3%; Score 113; DB 12; Length 1095;

Best Local Similarity 21.0%; Pred. No. 0.13; Indels 144; Gaps 19;
Matches 85; Conservative 54; Mismatches 121;

Qy 43 VLBDTKDILVAGPAGAEAIPIKFWLVVPCALIFMLI-----YAKLSNLSKQALFY 96
Db 677 ISRDYKN-----WKIILCLFIIPVCGFVSFRKKVPDKHKKLWY 717
Qy 97 AVG---TPFLI-----FFALPPTV-IYPLRDVLHPTFRDRLQALIPGLGLVA 142
Db 718 YVAFETSPFVFSNNVVFYIAFLFLFAVLLMDFSVHPPELV-----LYSLVF 767
Qy 143 IL-----RNMTEFAFYVLAELMGSVMLSL-MWGFANEITKIHAKRFYALFGIGANISL 196
Db 768 VLPCEDEVQRYVNVNFTDLM-NVMDTLGLFYFAGIVFRH-----SSNKS 815
Qy 197 LASGRAIWMASKRASVSEGVDPWGISRLIMAMTIVSGVLWASVWINKNVILDPREFY 256

Db 816 LYSGRVIFC-----LDYIIFTLRLIHIFTV----- 840
Qy 257 NPEMOKGKAGKAPKMMKDSFLYLDSPYLLTLTLVAVGICINLI----- 304
Db 841 -----SRNLGPKIIMLQRM-L-IDVFFFLFAVMMVAFGAROGILRQNEQRWIF 891
Qy 305 -EVTWKSQKL--QYPMNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKF-GMLTGALV 360
Db 892 RSVIYEPIYAMFGVPSVDVGTGYDPAHCTFTGNSKPLCVLDEHNLPRPEWITIPLV 951
Qy 361 TPVWVLLTGIVFALVIFRNQASGLVAMFGTTPMLAVVGAIQ 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVQ 978

RESULT 22

US-10-294-025-778
; Sequence 778, Application US/10294025
; Publication No. US20030185830A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Stolk, John A.
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C29
; CURRENT APPLICATION NUMBER: US/10/294,025
; NUMBER OF SEQ ID NOS: 1038
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 778
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-294-025-778

Query Match 4.3%; Score 113; DB 12; Length 1095;

Best Local Similarity 21.0%; Pred. No. 0.13; Indels 144; Gaps 19;
Matches 85; Conservative 54; Mismatches 121;

Qy 43 VLBDTKDILVAGPAGAEAIPIKFWLVVPCALIFMLI-----YAKLSNLSKQALFY 96
Db 677 ISRDYKN-----WKIILCLFIIPVCGFVSFRKKVPDKHKKLWY 717
Qy 97 AVG---TPFLI-----FFALPPTV-IYPLRDVLHPTFRDRLQALIPGLGLVA 142
Db 718 YVAFETSPFVFSNNVVFYIAFLFLFAVLLMDFSVHPPELV-----LYSLVF 767
Qy 143 IL-----RNMTEFAFYVLAELMGSVMLSL-MWGFANEITKIHAKRFYALFGIGANISL 196
Db 768 VLPCEDEVQRYVNVNFTDLM-NVMDTLGLFYFAGIVFRH-----SSNKS 815
Qy 197 LASGRAIWMASKRASVSEGVDPWGISRLIMAMTIVSGVLWASVWINKNVILDPREFY 256
Db 816 LYSGRVIFC-----LDYIIFTLRLIHIFTV----- 840
Qy 257 NPEMOKGKAGKAPKMMKDSFLYLDSPYLLTLTLVAVGICINLI----- 304
Db 841 -----SRNLGPKIIMLQRM-L-IDVFFFLFAVMMVAFGAROGILRQNEQRWIF 891
Qy 305 -EVTWKSQKL--QYPMNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKF-GMLTGALV 360
Db 892 RSVIYEPIYAMFGVPSVDVGTGYDPAHCTFTGNSKPLCVLDEHNLPRPEWITIPLV 951
Qy 361 TPVWVLLTGIVFALVIFRNQASGLVAMFGTTPMLAVVGAIQ 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVQ 978

RESULT 23

US-10-012-896-778
; Sequence 778, Application US/10012896
; Publication No. US20020183251A1

```
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yugu
APPLICANT: Kaios, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hedler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Bassols, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Fanger, Gary R.
APPLICANT: Wantanabe, Yoshihiro
APPLICANT: Meagher, Madeleine Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C27
CURRENT FILING DATE: 2001-12-10
NUMBER OF SEQ ID NOS: 1011
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 778
LENGTH: 1095
TYPE: PRT
ORGANISM: Homo sapiens
US-10-012-896-778

Query Match 4.3%; Score 113; DB 14; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.13;
Matches 85; Conservative 54; Mismatches 121; Indels 144; Gaps 19;

43 VLARDTKTLIVGAGSGAIAIPFIKFWLVPCALIFMLI-----YAKLSNLSKQALFY 96
Db 677 ISRDYKN-----WKIILCLFIPLVGGGFVSFRKKPVDKKKKLLWY 717

97 AVG---TPFLI-----FPALFPTV-IYPLRDVLHPTFADRLQAILPGLGLGVA 142
Db 718 YVAFSTSPFVFSNNVFIYAFLLFPAVYVLMDFHSVPHPELV-----LYSLVF 767

143 IL-----RNWTFAAFVYLAELMGSVMLSL-MFWGFANEITKIHEAKRFYALFGIGANISL 196
Db 768 VLFCEDEVQVYVGVNVTDLW-NVMDTGLGFYIAGIVFRLH-----SSNKSS 815

197 LASGRAIVMAKLRASVSEGVDPWGISRLMAMTIYSGVLWASVWINKVLTDRFY 256
Db 816 LYSGRVIFC-----LDYIIFTLRLIHFTV----- 840

257 NPEEMGKKGAKPKMMKDSFYLDSPYILLTLTVIAYGICINLI----- 304
Db 841 -----SNLGRKTIIMLQRM-LIDVFFFLFPAVMVAFGAROGILRQNEGRWIF 891

305 -EYTWKSQALKL--QYPMNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKE-GWLTGALV 360
Db 892 RSVIYEPLAMFGQVPSVDGTVDFAHCTFTGNEKSKPLCVELDEHNLPPEPWTITPLV 951

QY 361 TPWVUULTGIVFPAIVIFRNQASGIIVAMEGTTPLMLAVVGAIQ 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVO 978
```

RESULT 24
US-10-205-823-421

```
Sequence 421, Application US/10205823
Publication No. US20030108963A1
GENERAL INFORMATION:
APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Endege, Wilson O.
APPLICANT: Gannavarapu, Manjula
APPLICANT: Gorbacheva, Bella
APPLICANT: Hoersch, Sebastian
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Monsey, Angela M.
APPLICANT: Glatc, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Anderson, Dustin
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
FILE REFERENCE: MRI-044
CURRENT FILING DATE: 2002-07-25
CURRENT APPLICATION NUMBER: US/10/205, 823
PRIOR FILING DATE: 2001-07-25
PRIOR APPLICATION NUMBER: 60/307, 982
PRIOR FILING DATE: 2001-07-25
PRIOR APPLICATION NUMBER: 60/314, 356
PRIOR FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/325, 020
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: 60/341, 746
PRIOR FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: 60/362, 158
PRIOR FILING DATE: 2002-03-05
NUMBER OF SEQ ID NOS: 455
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 421
LENGTH: 1095
TYPE: PRT
ORGANISM: Homo sapiens
US-10-205-823-421

Query Match 4.3%; Score 113; DB 15; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.13;
Matches 85; Conservative 54; Mismatches 121; Indels 144; Gaps 19;

43 VLARDTKTLIVGAGSGAIAIPFIKFWLVPCALIFMLI-----YAKLSNLSKQALFY 96
Db 677 ISRDYKN-----WKIILCLFIPLVGGGFVSFRKKPVDKKKKLLWY 717

97 AVG---TPFLI-----FPALFPTV-IYPLRDVLHPTFADRLQAILPGLGLGVA 142
Db 718 YVAFSTSPFVFSNNVFIYAFLLFPAVYVLMDFHSVPHPELV-----LYSLVF 767

143 IL-----RNWTFAAFVYLAELMGSVMLSL-MFWGFANEITKIHEAKRFYALFGIGANISL 196
Db 768 VLFCEDEVQVYVGVNVTDLW-NVMDTGLGFYIAGIVFRLH-----SSNKSS 815

197 LASGRAIVMAKLRASVSEGVDPWGISRLMAMTIYSGVLWASVWINKVLTDRFY 256
Db 816 LYSGRVIFC-----LDYIIFTLRLIHFTV----- 840

257 NPEEMGKKGAKPKMMKDSFYLDSPYILLTLTVIAYGICINLI----- 304
Db 841 -----SNLGRKTIIMLQRM-LIDVFFFLFPAVMVAFGAROGILRQNEGRWIF 891

305 -EYTWKSQALKL--QYPMNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKE-GWLTGALV 360
Db 892 RSVIYEPLAMFGQVPSVDGTVDFAHCTFTGNEKSKPLCVELDEHNLPPEPWTITPLV 951

QY 361 TPWVUULTGIVFPAIVIFRNQASGIIVAMEGTTPLMLAVVGAIQ 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVO 978
```

RESULT 25
US-10-171-319-11

Sequence 11, Application US/10171319
Publication No. US20030157633A1
GENERAL INFORMATION:
APPLICANT: Arden Patapoutian
APPLICANT: Andrea Peler
APPLICANT: Peter McIntyre
APPLICANT: Stuart Bevan
APPLICANT: Chuansheng Song
APPLICANT: Pamosh Ganju
TITLE OF INVENTION: VANILLOID RECEPTOR-RELATED NUCLEIC ACIDS
FILE REFERENCE: 4-32048A
CURRENT APPLICATION NUMBER: US/10/171,319
CURRENT FILING DATE: 2002-10-24
PRIOR APPLICATION NUMBER: 60/297,835
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: 60/351,238
PRIOR FILING DATE: 2002-01-22
PRIOR APPLICATION NUMBER: 60/352,914
PRIOR FILING DATE: 2002-01-29
PRIOR APPLICATION NUMBER: 60/357,161
PRIOR FILING DATE: 2002-02-12
PRIOR APPLICATION NUMBER: 60/381,086
PRIOR FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: 60/381,739
PRIOR FILING DATE: 2002-05-16
NUMBER OF SEQ ID NOS: 114
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 1268
TYPE: PRT
ORGANISM: Homo sapiens
US-10-171-319-11

Query Match 4.3%; Score 113; DB 12; Length 1268;
Best Local Similarity 21.0%; Pred. No. 0.16;
Matches 85; Conservative 54; Mismatches 121; Indels 144; Gaps 19;

QY 43 VLRTKOTLIVGAPSGAEALPIFKMLVPCALIFMLI-----YAKLSNLSKQALFY 96
DB 850 ISRDTKN-----WKIILCLFPIIVGCGFVSFRKKPYDKHKKLWY 890
QY 97 AVG---TPFLI-----FFALPPTV-IYPLADVLHPTFEADRLQAILPPGLGLVA 142
DB 891 YVAFPTSPFVPSNVVYIYAFLLIFAYVLMDFHSVHPBELV-----LYSLVF 940
QY 143 IL-----RNTWFAFYVLAELWGSYMLSL-MFWGFRANITIKIHEAKRYVALFGIGANISL 196
DB 941 VLFCEVQWYVNGVNYFTDLW-NVMOTLGLFYFAGIVFRLH-----SSNKS 988
QY 197 LASGRAIYMAKSLRASVSEGVDPWGISIRFLMAMTIVSGVLVMSYMWINKVLTDPFRY 256
DB 989 LYSGRVIFC-----LDYIIFTLRLIHIFTV----- 1013
QY 257 NPEMQKKKAKPKMNKDSFLYLDSPYILLLLVIAVIGICINLI----- 304
DB 1014 -----SRNIGPKITIMLQRM-LIDVFFFLFPAVMAAGVAAQGLRQNEGRWIF 1064
QY 305 -EVTWKSQKL--QYNNNDYSEFWGNFSFWTGVSVLIMLVGNAVIRKF-GWLTGALV 360
DB 1065 RSVIYEPFLAMGCVPSVDGTYTYPACHTFGNSKPLCYELDBHNLPRFPWITITPLV 1124
QY 361 TPVWVLLTGIVFPALVIRNOASGVAMFGTTPLMLAVVGAIQ 404
DB 1125 C-IYVLTSTNILLVNL-----LVAMFGYT-----VGVVQ 1151

RESULT 26
US-09-920-653-3
Sequence 3, Application US/09920653
Patent No. US20020104113A1
GENERAL INFORMATION:
APPLICANT: Japan as Represented by Director General of Okazaki National Research

APPLICANT: Instituees
TITLE OF INVENTION: Nav2 channel gene-deficient non-human animals
FILE REFERENCE: J2001P059
CURRENT APPLICATION NUMBER: US/09/920,653
CURRENT FILING DATE: 2001-10-16
PRIOR APPLICATION NUMBER: JP 2000/237320
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: JP 2000/241637
PRIOR FILING DATE: 2000-08-09
PRIOR APPLICATION NUMBER: JP 2001/222263
PRIOR FILING DATE: 2001-07-23
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 1681
TYPE: PRT
ORGANISM: Mus musculus
US-09-920-653-3

Query Match 4.3%; Score 113; DB 10; Length 1681;
Best Local Similarity 21.4%; Pred. No. 0.23; Indels 106; Gaps 18;
Matches 78; Conservative 56; Mismatches 124;

QY 3 KTEEKPEGRKSRPFLPIHTHELKKVLPMLMFCITFNVTYVLRDTKOTLIVGAPSGAEA 62
DB 1231 KPTPRPNKKGGLFDLVTHRVFVILIL-----ICFOATTIMIQD-----EQ 1275
QY 63 IPIFK--FWL-----VPCAI-----IMLYAKLSNLSKQAL 94
DB 1276 SPQWETAIFWVNSIFVNLFTLECLIKLAPRCHYFSAVNVHDVVVIFSTGLLPVTI 1335
QY 95 FYAVGTPELIFFAFPYIVYELRDVLHPTFEADRLQAILPPGLGLVAIIRNTWFAFYV 154
DB 1336 GQYVPPSLVOLILSNVYIHLRPGKPKVPHD-----LMDPLIALALL-NISLLIFLV 1390
QY 155 LAELWGSYMLMFWGFRANITIKIHEAKRYVALFGIGANISLILASGRAIYMAKSLRASVS 214
DB 1391 -----NF-----IYAFGM-YNFAYV-----KKEAGIN 1412
QY 215 E-GVDWGISLRLMAMTIVSG---LVLMASYMWINKVLTDPFRYNPEMQKKKGA 268
DB 1413 DVSNFEETFGSMCLFQVITTFSGWDMIDAIENSQWSD---CPDDKINPQYKGCQS 1468
QY 269 KPKNMKDSFLYDRSPYILLTLVTAAGICINLIVTWKSQKLQYNNNDYSEFWGN 328
DB 1469 -PSVGT--SYFV---SYLISWLIYVMTIVLIMEFLSLSPQKSKRTLSDDPRF--- 1518
QY 329 FSPW 332
DB 1519 FRVW 1522

RESULT 27
US-09-738-626-6914
Sequence 6914, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIJO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: PatentIn ver. 3.0
 ; SEQ ID NO: 6914
 ; LENGTH: 1083
 ; TYPE: PRF
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-6914

Query Match 4.2%; Score 111.5; DB 10; Length 1083;
 Best Local Similarity 19.0%; Pred. No. 0.18;
 Matches 114; Conservative 80; Mismatches 170; Indels 237; Gaps 28;

QY 12 LRSFLMPLTHLKLKLVLMF--CITFNTVLRDTOTLIVAGPSGAEAIPIKFW 69
 DB 105 LLSLVIPVLTAKEDAGSGFRLTLSTVLGGVTLISIGAP----- 151
 QY 70 LVVPCALIFMLYAKLS-----NLSKALFYAVGPELIFALPFTVYPL--RDVLRP 122
 DB 152 -----LITRMMLSSGQVNVWSTAPAYWL-LPQIFYGLEFMAVINTREVEKP 201
 QY 123 TEFADRLQALPPGLIGLVAIILRMTEFAFYVLAELMGSVMLSMFEGFANEITKHEAK 182
 DB 202 GAMAPVNNVITLTVLGVVWLP-----ARLHPHE 231
 QY 183 RF-----YALFGIGANISLASGRAIYVASKIRASVSEGVDP--WGISLRLL--MA 229
 DB 232 QVGIRDPQIIFLGVGTLLGVA--OCLIMIPLYRRA--GIDMRPLMGIDARLKQFGGWA 286
 QY 230 MTIV-----SGLVMASYWMINKVLLD--PRFN 257
 DB 287 NALIVYVAISQFGYIITRIASIDADAPFIYQOHMLQVPEYGIIGVTLTALMERLSR 346
 QY 258 -----PEMOKKKKAKPKKMMKDSPLYLRSFYILLTLTLLVI--AYGICINLIE 305
 DB 347 NAADDDBRVDLQDLSK-----LTFIALIPVFTAGV----- 383
 QY 306 VTMKSQKLQYNNND--YSEPMGN-----FSPFTGVSVLIMLVGVGNVIRKP 352
 DB 384 -----PIANGLFAYGQFDANAANILGWTLSFSAFTLIPYALVTLHLRVEYARE 432
 QY 353 GMLTGAIVTPVWVLLTGIVFALVIFRNASGLVAMEGTPMLAVVCAIQNI----- 406
 DB 433 VM-----TPTFT-TAGITATKVL-----SLAPLSSSPERVVLLGANGFSFITGA 480
 QY 407 -----LSKSTKYALFDSKEMAVIPLDQEKYKKAIDVVA--RFG 447
 DB 481 VIGAVLNNKGLGLGMRSLAKTSLWL-----GSAVGAAMAAALMG 521
 QY 448 KSGGALIOGGLIVIGSIGAMTPYLAIVILLFTIIMVLSATKLKFLAQSALKE-QEVA 506
 DB 522 WLIGAVGDFLLGLTSSVGYLL-YLAVLGVFIFV-----TGVIVSRSGLPVQNLG 572
 QY 507 Q 507
 DB 573 Q 573

RESULT 28
 US-09-738-626-4455
 ; Sequence 4455, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAOKO

; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1998-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: PatentIn ver. 3.0
 ; SEQ ID NO: 4455
 ; LENGTH: 741
 ; TYPE: PRF
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-4455

Query Match 4.2%; Score 111; DB 10; Length 741;
 Best Local Similarity 18.9%; Pred. No. 0.12;
 Matches 108; Conservative 83; Mismatches 205; Indels 176; Gaps 24;

QY 27 VLMFMEFCIT-----FNYTVLRDTOTLIVAGPSGAEAIPIKFWLV 72
 DB 204 LIPSTIMGLSTIIAVALMLISPEVFTISASRQTRTALLASQ-----ATPRHIRMVAVL 259
 QY 73 PCALIFMLYAKLSNLSKQALF-----YAVGTEPLI--FFAL-----FP 110
 DB 260 TYGLFAGLVGASIGLVLOIGIYGWKKTYBPFSITLTVLVGVWALAIASITAAFLP 319
 QY 111 TYIPLRDVLRHT--EPADRLOALIPGLIGLVAIILRMTEFAFYVLAELMGSVMLSMF 168
 DB 320 AVFVSRSSIIINGIYIGSIDKIIIRWSPRMILGPIVLAIAVIALFIDGGE-WGVVQKQCF 378
 QY 169 WGFANEIRKIHAKFYVLFPGIGANISLASGRAIYVA-----SKURASVS 214
 DB 379 -----IAAVIALPASVPALVLMALGRPLGTLFKLATRDMLRSM- 416
 QY 215 EGVDFWEGISLRILNMTIV--SGLVMASYWMINKVLLDPRFYPEEMOKKKKAKPKM 272
 DB 417 HSIPALGALAVIMLGTMTGTATQASDEATASVPEAVFLRGDTQIPIGMOKIDV 476
 QY 273 ---NNKDSFLYD---RSPYILLTLL---VIAYGICINLIEVWKSQKLQYNN--- 318
 DB 477 YGDHNGFGIYELDVDFYSANVVPALTSFEGEPVIAIPKILDMFGV--HEQADIVAPSTVN 534
 QY 319 ---NDYSEFMGNFSWTVGVSVLIMF---VGVNVIRKFGMLT--GALVTP----- 362
 DB 535 SGLQRYAIVPDDEYMLDTAVALPLPVLSHVLSPETFEIIGQTEFLGTIVLPOELDQ 594
 QY 363 -----VMVLLTGIVFALVIF-----RNOASGLVAMEGTP- 393
 DB 595 VQAINRSDAFPHSHDGHNSLASSAALTRAVAIIVSVLIVANRKLQOHALAL- GATPG 653
 QY 394 -----LMLAVVGAIONILSKSTKYALFDSKEMAVIPLDQEKYKKAIDVVA 444
 DB 654 TIYKVNALNALMLLVGIMGVSGVIMALLTGTIDEI-----VDGAIL 697
 QY 445 RFGSGGALLIOGGLIV-----ICSGTGM 468
 DB 698 NYGTLEHMLPWLPLVLSLVVAPLVCAGVIGAI 729

RESULT 29
 US-09-759-143-780
 ; Sequence 780, Application US/09759143
 ; Patent No. US20020022248A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiaengchun
 ; APPLICANT: Dillon, Davin C.

;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;; FILE REFERENCE: 210121.534C1
;; CURRENT APPLICATION NUMBER: US/09/822.827
;; CURRENT FILING DATE: 2001-03-28
;; NUMBER OF SEQ ID NOS: 982
;; SOFTWARE: FASTSEQ for Windows Version 3.0
;; SEQ ID NO 780
;; LENGTH: 1095
;; TYPE: PRF
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: VARIANT
;; LOCATION: (1)...(1095)
;; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-822-827-780

Query Match 4.2%; Score 111; DB 9; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.2;
Matches 85; Conservative 53; Mismatches 122; Indels 144; Gaps 19;

Qy 43 VLRTKDTLIVGARGSGAEALPIKFWLVPCAIIFMLI-----YAKLSNLSKQALFY 96
Db 677 ISRDTKN-----WKIILCLFIPLVCGGFVSFRKKPVDKHKKILMY 717

Qy 97 AVG---TPFLI-----FPALFPTV-IYPLRDVLPTEFADRLQALIPGLGLIVA 142
Db 718 YVAFSTSPFVFSNNVFIYAFLLFAVLLMDPHSHVPELV-----LYSLVF 767

Qy 143 IL-----RNMTFAAFVYLAELMGSVMLSL-MFWGFANEITKIHAKRYALFGIGANISL 196
Db 768 VLFCDERQWYVNGVNYFTDLM-NVMDTLGLFYIAGIVFRHL-----SSNKSS 815

Qy 197 LASGRAIVMAASKLRASVSEGVDPWGISLRLLMAMTIYSGVLVMSYWMINKVLTDRPFY 256
Db 816 LYSGRVIFC-----LDYIIFTLRLIHIFTV----- 840

Qy 257 NPEMQGKKKAKPRKMKDSFLYIDRSPIYLLTLTVIAGICINLI----- 304
Db 841 -----SRNLGPKIIMLQML-IDVFFFLFPAKMMVAFGVARQGLLRONEQRWEIF 891

Qy 305 -EYTWKSQQLK--QYPMNDYSEFMGNFSFWTGVSVLIMLVGQNVIRKF-GMLTGALV 360
Db 892 RSVIYBFLAMFGQVPSVDGDTYDFACHTFTGNSKPLCVBLDEHNLPRPEWITIPLV 951

Qy 361 TPVAVLLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVGAIQ 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVQ 978

RESULT 32
US-09-895-793-780
; Sequence 780, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.

;; APPLICANT: Houghton, Raymond L.
;; APPLICANT: Vinals de Bassols, Carlota
;; APPLICANT: Foy, Teresa
;; APPLICANT: Fanger, Gary R.
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;; FILE REFERENCE: 210121.534C2
;; CURRENT APPLICATION NUMBER: US/09/895.793
;; CURRENT FILING DATE: 2001-06-29
;; NUMBER OF SEQ ID NOS: 982
;; SOFTWARE: FASTSEQ for Windows Version 3.0
;; SEQ ID NO 780
;; LENGTH: 1095
;; TYPE: PRF
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: VARIANT
;; LOCATION: (1)...(1095)
;; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-895-793-780

Query Match 4.2%; Score 111; DB 10; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.2;
Matches 85; Conservative 53; Mismatches 122; Indels 144; Gaps 19;

Qy 43 VLRTKDTLIVGARGSGAEALPIKFWLVPCAIIFMLI-----YAKLSNLSKQALFY 96
Db 677 ISRDTKN-----WKIILCLFIPLVCGGFVSFRKKPVDKHKKILMY 717

Qy 97 AVG---TPFLI-----FPALFPTV-IYPLRDVLPTEFADRLQALIPGLGLIVA 142
Db 718 YVAFSTSPFVFSNNVFIYAFLLFAVLLMDPHSHVPELV-----LYSLVF 767

Qy 143 IL-----RNMTFAAFVYLAELMGSVMLSL-MFWGFANEITKIHAKRYALFGIGANISL 196
Db 768 VLFCDERQWYVNGVNYFTDLM-NVMDTLGLFYIAGIVFRHL-----SSNKSS 815

Qy 197 LASGRAIVMAASKLRASVSEGVDPWGISLRLLMAMTIYSGVLVMSYWMINKVLTDRPFY 256
Db 816 LYSGRVIFC-----LDYIIFTLRLIHIFTV----- 840

Qy 257 NPEMQGKKKAKPRKMKDSFLYIDRSPIYLLTLTVIAGICINLI----- 304
Db 841 -----SRNLGPKIIMLQML-IDVFFFLFPAKMMVAFGVARQGLLRONEQRWEIF 891

Qy 305 -EYTWKSQQLK--QYPMNDYSEFMGNFSFWTGVSVLIMLVGQNVIRKF-GMLTGALV 360
Db 892 RSVIYBFLAMFGQVPSVDGDTYDFACHTFTGNSKPLCVBLDEHNLPRPEWITIPLV 951

Qy 361 TPVAVLLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVGAIQ 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVQ 978

RESULT 33
US-09-895-814-780
; Sequence 780, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Heppler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hurai, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Bassols, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Fanger, Gary R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C26
CURRENT APPLICATION NUMBER: US/09/895,814
CURRENT FILING DATE: 2001-06-29
NUMBER OF SEQ ID NOS: 990
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 780
LENGTH: 1095
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(1095)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-895-814-780

Query/Match 4.2%; Score 111; DB 10; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.2; Indels 144; Gaps 19;
Matches 85; Conservative 53; Mismatches 122;

QY 43 VLKPTKTLIVGAPSGAEAFPIKFWLVVPCALIFMLI-----YAKLSNLSKQALFY 96
DB 677 ISRDTKN-----WKIILCLFIPLVCGCFVSFRKKPVDKHKKLLMY 717
QY 97 AVG---TFPLI-----FFALPPTV-IYPLRDVLHPTFADRLQAILPGLGLVA 142
DB 718 YVAFSTSPFVFSNVVFFIYAFLLFAYVLLMDPFSVHPPELV-----LYSLVF 767
QY 143 IL-----RNMTFAFYVLAELMGSVMLS-MWGFANETIKHEAKRYALFGICANISL 196
DB 768 VLFCEVRYQWYVGVNFTDLM-NVMDTLGLFTIAGIVFRHL-----SSNKS 815
QY 197 LASGRAIVWASKLRAVSSEGVDPWGISRLMAMTIVSGVLVMAWYMWINKNVLTDPFV 256
DB 816 LYSGRVIFC-----LDYIIFTLRLHIHFTV----- 840
QY 257 NPEMQKKKGAQKPKMMKDSFLYDRSPYILTLTVIYAGICINLI----- 304
DB 841 -----SRNLGPKIIMLQRMIL-IDVFFFLFLFAKMMVAFGVARQGLRQNEQRMWIF 891
QY 305 -EYTWKSQTL--QYPMNDYSEFMGNFSFWTGVSVLIMLFVGNAVIRKF-GWLTGALV 360
DB 892 RSVIYEPYLAHFGVPSVDGTYDPAHCTFTGNEKFLCVELDEHNLPRPEWITPLV 951
QY 361 TPVWVLLTGIVFALVIFRNQASGLVAMFGTTPMLAVVGAIQ 404
DB 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVQ 978

RESULT 34
US-10-144-678A-780
Sequence 780, Application US/10144678A
Publication No. US20030157089A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yugu
APPLICANT: Henderson, Robert A.
APPLICANT: Fanger, Gary R.
APPLICANT: Kalos, Michael D.
APPLICANT: Stoik, John A.

APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Derrick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A. W.
APPLICANT: Heppler, William T.
APPLICANT: Hurai, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals y de Bassols, Carlota
APPLICANT: Foy, Teresa M.
APPLICANT: Matanabe, Yoshihiro
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C28
CURRENT APPLICATION NUMBER: US/10/144,678A
CURRENT FILING DATE: 2002-08-12
NUMBER OF SEQ ID NOS: 1033
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 780
LENGTH: 1095
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: 867
OTHER INFORMATION: Xaa = Any Amino Acid
US-10-144-678A-780

Query/Match 4.2%; Score 111; DB 12; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.2; Indels 144; Gaps 19;
Matches 85; Conservative 53; Mismatches 122;

QY 43 VLKPTKTLIVGAPSGAEAFPIKFWLVVPCALIFMLI-----YAKLSNLSKQALFY 96
DB 677 ISRDTKN-----WKIILCLFIPLVCGCFVSFRKKPVDKHKKLLMY 717
QY 97 AVG---TFPLI-----FFALPPTV-IYPLRDVLHPTFADRLQAILPGLGLVA 142
DB 718 YVAFSTSPFVFSNVVFFIYAFLLFAYVLLMDPFSVHPPELV-----LYSLVF 767
QY 143 IL-----RNMTFAFYVLAELMGSVMLS-MWGFANETIKHEAKRYALFGICANISL 196
DB 768 VLFCEVRYQWYVGVNFTDLM-NVMDTLGLFTIAGIVFRHL-----SSNKS 815
QY 197 LASGRAIVWASKLRAVSSEGVDPWGISRLMAMTIVSGVLVMAWYMWINKNVLTDPFV 256
DB 816 LYSGRVIFC-----LDYIIFTLRLHIHFTV----- 840
QY 257 NPEMQKKKGAQKPKMMKDSFLYDRSPYILTLTVIYAGICINLI----- 304
DB 841 -----SRNLGPKIIMLQRMIL-IDVFFFLFLFAKMMVAFGVARQGLRQNEQRMWIF 891
QY 305 -EYTWKSQTL--QYPMNDYSEFMGNFSFWTGVSVLIMLFVGNAVIRKF-GWLTGALV 360
DB 892 RSVIYEPYLAHFGVPSVDGTYDPAHCTFTGNEKFLCVELDEHNLPRPEWITPLV 951
QY 361 TPVWVLLTGIVFALVIFRNQASGLVAMFGTTPMLAVVGAIQ 404
DB 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVQ 978

RESULT 35
US-10-294-025-780
Sequence 780, Application US/10294025
Publication No. US20030185830A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Stoik, John A.
APPLICANT: Kalos, Michael D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND


```

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10467
; LENGTH: 425
; TYPE: PR
; ORGANISM: Escherichia coli
US-09-815-242-10467

```

```

Query Match      4.2%; Score 110.5; DB 9; Length 425;
Best Local Similarity 19.7%; Pred. No. 0.063;
Matches 93; Conservative 68; Mismatches 139; Indels 173; Gaps 24;

```

```

QY 84 KLSNLSKQALFPA-----VGTPELFFALFPYIVLRDVLPHTERADRLQALIPGL 137
DB 27 KQVNPQKRALFSAMLGIVDFDPDMIF-----YLIHIKADLGITD---IQAT---L 74
QY 138 LGLVALIRNMTFAAFY-VLAELMGSVMLSLMFWG-FANEITKIHAKRFYALFGIGANIS 195
DB 75 IGTVAFIARPIGGFFGAMADKYG--RKPMMMAIIFYSGT-----GLSGIATNLY 124
QY 196 LLAGRAIV--WASKLRASVSEGVDPW-----GISRLIMAMTIVS--- 234
DB 125 MLAVCRFIVGLGMSGEYACASTYAVESWPKNLQSKASAFVSGSVGNIIAQTIIPOFAE 184
QY 235 -----GLVMASTWINKNVLTDRFPNPEEMQKKGKAKPKKMMKDSFLYDPS 284
DB 185 VYGMNSFFIGLLPVLVLMIRKSA-----PE-----SQEWIEDKXDKSTFLSVFRK 232
QY 285 PYILLTLTLVIAVIGICINLIEVTWKSQKLQYPMNDYSEFMGNFSFMTGVSVLIML-- 342
DB 233 PH---LSISMIVFLVCFCLFGANW-----PINGLEPSYLDANGVNTVISTLMTING 281
QY 343 -----FVGNVIRKFGMLTGAIVTPVMTL-----LTGIVFALVIFR 379
DB 282 LGLTNGTIFFGFVDKIGVKAFAVVG-LITSFIFLCPFFISVKNSSLIIGLCFLGIMFTN 340
QY 380 NQASGLVAMFGTTPLMLAVVGAIONILSKSTKVALFDSSTKEMAYIPLDQEQKYGKAKI 439
DB 341 LGIAGLVPRF-----IYD-----YFP-----TKLRG--- 361
QY 440 DVVAARFGSGALIQGLLVIGSIGAM-TPYLAV-----ILLFIITAI 482
DB 362 -----LGTGLIYNLGATGMAAPVLAITYISGYGGLSFLIVTV 400

```

```

RESULT 38
US-10-287-274-428
; Sequence 428, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THEREOF
; FILE REFERENCE: ELITRA, 0080V1
; CURRENT APPLICATION NUMBER: US/10/287,274
; PRIOR FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09

```

```

; PRIOR APPLICATION NUMBER: US 09/711164
; PRIOR FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 428
; LENGTH: 425
; TYPE: PR
; ORGANISM: Escherichia coli
US-10-287-274-428

```

```

Query Match      4.2%; Score 110.5; DB 12; Length 425;
Best Local Similarity 19.7%; Pred. No. 0.063;
Matches 93; Conservative 68; Mismatches 139; Indels 173; Gaps 24;

```

```

QY 84 KLSNLSKQALFPA-----VGTPELFFALFPYIVLRDVLPHTERADRLQALIPGL 137
DB 27 KQVNPQKRALFSAMLGIVDFDPDMIF-----YLIHIKADLGITD---IQAT---L 74
QY 138 LGLVALIRNMTFAAFY-VLAELMGSVMLSLMFWG-FANEITKIHAKRFYALFGIGANIS 195
DB 75 IGTVAFIARPIGGFFGAMADKYG--RKPMMMAIIFYSGT-----GLSGIATNLY 124
QY 196 LLAGRAIV--WASKLRASVSEGVDPW-----GISRLIMAMTIVS--- 234
DB 125 MLAVCRFIVGLGMSGEYACASTYAVESWPKNLQSKASAFVSGSVGNIIAQTIIPOFAE 184
QY 235 -----GLVMASTWINKNVLTDRFPNPEEMQKKGKAKPKKMMKDSFLYDPS 284
DB 185 VYGMNSFFIGLLPVLVLMIRKSA-----PE-----SQEWIEDKXDKSTFLSVFRK 232
QY 285 PYILLTLTLVIAVIGICINLIEVTWKSQKLQYPMNDYSEFMGNFSFMTGVSVLIML-- 342
DB 233 PH---LSISMIVFLVCFCLFGANW-----PINGLEPSYLDANGVNTVISTLMTING 281
QY 343 -----FVGNVIRKFGMLTGAIVTPVMTL-----LTGIVFALVIFR 379
DB 282 LGLTNGTIFFGFVDKIGVKAFAVVG-LITSFIFLCPFFISVKNSSLIIGLCFLGIMFTN 340
QY 380 NQASGLVAMFGTTPLMLAVVGAIONILSKSTKVALFDSSTKEMAYIPLDQEQKYGKAKI 439
DB 341 LGIAGLVPRF-----IYD-----YFP-----TKLRG--- 361
QY 440 DVVAARFGSGALIQGLLVIGSIGAM-TPYLAV-----ILLFIITAI 482
DB 362 -----LGTGLIYNLGATGMAAPVLAITYISGYGGLSFLIVTV 400

```

```

RESULT 39
US-09-738-626-5741
; Sequence 5741, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: KENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059

```

SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5741
LENGTH: 476
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5741

Query Match 4.2%; Score 110.5; DB 10; Length 476;
Best Local Similarity 21.6%; Pred. No. 0.073; Indels 79; Gaps 14;
Matches 55; Conservative 44; Mismatches 77;

QY 209 LRASVSECVD-----PWGISRLMAMTIVSGVLMAASYWMINKNVLTDRPFYNE--E 260
DB 205 LIASLASQMDVNIWTWNAV-----AIVPGVALIVPMVYVKI-----YPELKD 251
QY 261 MOKGKKGAKPKMNMDSFLYIDR---SPYILILT-----LVIA-----YGCIMLI 304
DB 252 TPEVKQASDRIKQIGFTYGEKVLGTFVVLILMTGDLVLGISATTAFVGVILLV 311
QY 305 E--VTWKS--OLKIQPMNNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKFG--WITG 357
DB 312 AHVLTWEDIIEKJAMDITM-----VWPAV-----LYMMATALSQYGFIAWISE 354
QY 358 ALVTPV-----MVLITGIVFPALVIFRNOASGLVAMF-----GTPPLMLAV 399
DB 355 VIASSIGGMNVVALVVLVLYFPFSHYFPASATAHISMYLAFLGAALAIAGAPPLMALV 414
QY 400 VCAIQNIISKSTKYA 414
DB 415 LAYTSNLFSSLTQYS 429

RESULT 40

US-10-032-585-7923
Sequence 7923, Application US/10032585
Publication No. US20030180953A1
GENERAL INFORMATION:
APPLICANT: Terry, Roemer D.
APPLICANT: Bo, Jiang
APPLICANT: Charles, Boone
APPLICANT: Howard, Bussey
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
FILE REFERENCE: 10182-005-999
CURRENT APPLICATION NUMBER: US/10/032,585
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 8000
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7923
LENGTH: 927
TYPE: PRT
ORGANISM: Candida albicans
US-10-032-585-7923

Query Match 4.2%; Score 110; DB 12; Length 927;
Best Local Similarity 20.9%; Pred. No. 0.2; Indels 160; Gaps 24;
Matches 90; Conservative 60; Mismatches 121;

QY 66 IKFELVVPICAIIFMLIAKLSNI--LSKQALFVAVGPFLIFPALFTVIYPLRDVLHPT 123
DB 397 IIFW-CIPVAV-----GAISINIVLTDR-----VFLKFIILKMPDIVINGVITGLLPV 443
QY 124 EFADRLGAILPRLGLLVALIRNWTFAFVYLAELMGVMLS-----LMFGFANEITKI 178
DB 444 VALAILMSLVDPFI-----KM-----MGKISGRLLTIQVYESYCQSYFAFQVVNV 488
QY 179 HEAKRYALFGIGANISILAS-----GRAIVWASKLRAVSEGVDPWGISRLIMAMTI 232
DB 489 -----FLAIALGSSAAVAATQIVQNPGEAL--QKLSSEFPKSVN--FYYSYLCLEGITI 538
QY 233 VSGVLAM-----ASYWINKNVLTDRFYNPEMOKGKKGAAPKRNMK 275
DB 539 SSGVTLQIVALLSHILGRILIDGTPRAKWTMRNTLIGQPAY----- 578

QY 276 DSFLYLDSPYIILTLTLVIAVGICINLIEVTWKSQKLQYPMNNDYSEFMGNFSFWTGV 335
DB 579 -STIX-----PQFQTLVIALSYVIAPI-----LG-----FTAI 608
QY 336 VSVLI-----MLFV-----GGNVIRK--FGWLTGALVTPVWVLTGIVEFA----- 374
DB 609 AFILFYFAYIYTMIFLBRPSTVDARGTNVYKSLFQFTGLFLAQLM--TALFVFSKWA 666
QY 375 -----LVIFRNOASGLVAMGTTPLMLAVVGAIQNIILSKSTKYALPDSIKENAYIPLD 428
DB 667 CVALEGVIVVVTIARLMMKKFLPLVDVAVISAI-----KYAAGDPT--YSYPIHD 716
QY 429 Q---EOKVKGK 436
DB 717 QGLKEIVBGR 727

Search completed: November 25, 2003, 10:21:55
Job time : 59 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:11:17 ; Search time 64 Seconds
(without alignments)
2076.517 Million cell updates/sec

Title: US-09-869-433-2
Perfect score: 2530
Sequence: 1 MTKREKPFGLRSLFPIH.....AQSLKGEVNAQEDSAPASS 515

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :

SPTREMBL 23: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_todent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriophage: *
17: sp_archaeal: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|-----------|---------------------|
| 1 | 2144 | 81.5 | 528 | 2 Q9SEV3 | Q9SEV3 chlamydia t |
| 2 | 1358.5 | 51.7 | 644 | 10 Q9AS91 | Q9AS91 oryza sativ |
| 3 | 1302.5 | 49.5 | 671 | 8 Q8WH67 | Q8WH67 galidieria s |
| 4 | 1302 | 49.5 | 588 | 8 Q8MFR8 | Q8MFR8 citrus hybr |
| 5 | 1076 | 40.9 | 498 | 16 Q9ZJ16 | Q9ZJ16 rickettsia |
| 6 | 1020 | 38.8 | 514 | 2 Q8KPN2 | Q8KPN2 holospora o |
| 7 | 976.5 | 37.1 | 501 | 16 Q9ZHP9 | Q9ZHP9 rickettsia |
| 8 | 960.5 | 36.5 | 540 | 2 Q9X9D2 | Q9X9D2 chlamydia t |
| 9 | 907.5 | 34.5 | 499 | 16 Q9ZG15 | Q9ZG15 rickettsia |
| 10 | 907.5 | 34.5 | 511 | 16 Q9ZHV4 | Q9ZHV4 rickettsia |
| 11 | 827 | 31.4 | 507 | 16 Q9ZJ98 | Q9ZJ98 rickettsia |
| 12 | 366 | 13.9 | 543 | 5 Q8SUG0 | Q8SUG0 encephalito |
| 13 | 351 | 13.3 | 553 | 5 Q8SUF9 | Q8SUF9 encephalito |
| 14 | 338.5 | 12.9 | 559 | 5 Q8SRA2 | Q8SRA2 encephalito |
| 15 | 325 | 12.4 | 536 | 5 Q8SUG7 | Q8SUG7 encephalito |
| 16 | 289.5 | 11.0 | 105 | 10 Q945F3 | Q945F3 medicago sa |

| | | | | | |
|----|-------|-----|------|-----------|--------------------|
| 17 | 229.5 | 8.7 | 441 | 16 Q9PCP0 | Q9PCP0 xylella fas |
| 18 | 203.5 | 7.7 | 454 | 16 Q8P6U3 | Q8P6U3 xanthomonas |
| 19 | 197 | 7.5 | 496 | 10 Q9FLC3 | Q9FLC3 xanthomonas |
| 20 | 188.5 | 7.2 | 453 | 16 Q8P142 | Q8P142 xanthomonas |
| 21 | 173.5 | 6.6 | 528 | 16 Q84237 | Q84237 chlamydia t |
| 22 | 162 | 6.2 | 538 | 16 Q8NU46 | Q8NU46 corynebacte |
| 23 | 161 | 6.1 | 918 | 16 Q9PKG0 | Q9PKG0 chlamydia m |
| 24 | 158.5 | 6.0 | 925 | 16 Q9ZAP5 | Q9ZAP5 chlamydia p |
| 25 | 157.5 | 6.0 | 925 | 16 Q9J516 | Q9J516 chlamydia p |
| 26 | 151 | 5.7 | 1002 | 16 Q8YND4 | Q8YND4 anabaena sp |
| 27 | 147 | 5.6 | 533 | 17 Q8TW21 | Q8TW21 methanosaic |
| 28 | 145 | 5.5 | 540 | 16 Q9KSH7 | Q9KSH7 vibrio chol |
| 29 | 144 | 5.5 | 660 | 16 Q9JUS7 | Q9JUS7 neisseria m |
| 30 | 144 | 5.5 | 675 | 16 Q9JZ61 | Q9JZ61 neisseria m |
| 31 | 141 | 5.4 | 563 | 16 Q8PUB5 | Q8PUB5 corynebacte |
| 32 | 140 | 5.3 | 474 | 16 Q9KFD0 | Q9KFD0 bacillus ha |
| 33 | 140 | 5.3 | 475 | 2 Q9Z9P3 | Q9Z9P3 bacillus ha |
| 34 | 137 | 5.2 | 536 | 8 Q20628 | Q20628 cryptocodi |
| 35 | 137 | 5.2 | 577 | 8 Q94Y38 | Q94Y38 cryptocodi |
| 36 | 136.5 | 5.2 | 498 | 2 Q33750 | Q33750 synechococc |
| 37 | 135.5 | 5.2 | 465 | 16 Q8FHZ0 | Q8FHZ0 escherichia |
| 38 | 135 | 5.1 | 441 | 16 Q916P7 | Q916P7 pseudomonas |
| 39 | 134.5 | 5.1 | 404 | 16 Q8G4F6 | Q8G4F6 bifidobacte |
| 40 | 134 | 5.1 | 976 | 17 Q9V250 | Q9V250 pyrococcus |
| 41 | 133.5 | 5.1 | 516 | 16 Q8FPUL | Q8FPUL corynebacte |
| 42 | 132.5 | 5.0 | 526 | 2 Q8MID5 | Q8MID5 chaetophae |
| 43 | 132 | 5.0 | 552 | 2 Q9APX1 | Q9APX1 pseudomonas |
| 44 | 131.5 | 5.0 | 463 | 16 Q8XDD0 | Q8XDD0 escherichia |
| 45 | 131.5 | 5.0 | 567 | 17 Q97VC8 | Q97VC8 sulfolobus |
| 46 | 131 | 5.0 | 434 | 16 Q8UDN2 | Q8UDN2 agrobacteri |
| 47 | 130.5 | 5.0 | 420 | 2 Q46305 | Q46305 clostridium |
| 48 | 130.5 | 5.0 | 976 | 17 Q74088 | Q74088 pyrococcus |
| 49 | 128.5 | 4.9 | 401 | 16 Q8ELQ1 | Q8ELQ1 oceanobacil |
| 50 | 128.5 | 4.9 | 446 | 16 Q8ZBT6 | Q8ZBT6 yersinia pe |
| 51 | 129.5 | 4.9 | 581 | 16 Q98Q22 | Q98Q22 mycoplasma |
| 52 | 129 | 4.9 | 484 | 8 Q47458 | Q47458 cryptocodi |
| 53 | 129 | 4.9 | 721 | 10 Q9FRN4 | Q9FRN4 oryza sativ |
| 54 | 128.5 | 4.9 | 435 | 16 Q99WS1 | Q99WS1 staphylococ |
| 55 | 128.5 | 4.9 | 665 | 16 Q9A7F1 | Q9A7F1 caulobacter |
| 56 | 128 | 4.9 | 462 | 17 Q97XMO | Q97XMO sulfolobus |
| 57 | 127.5 | 4.8 | 496 | 5 Q19676 | Q19676 caenorhabdi |
| 58 | 127 | 4.8 | 397 | 2 Q84999 | Q84999 streptococc |
| 59 | 127 | 4.8 | 462 | 16 Q8X9Y3 | Q8X9Y3 escherichia |
| 60 | 126.5 | 4.8 | 999 | 2 Q93JY2 | Q93JY2 errinia chr |
| 61 | 126 | 4.8 | 397 | 2 Q9R9Z5 | Q9R9Z5 streptococ |
| 62 | 126 | 4.8 | 513 | 10 Q94EC4 | Q94EC4 oryza sativ |
| 63 | 126 | 4.8 | 548 | 3 Q9P3K6 | Q9P3K6 neurospora |
| 64 | 125.5 | 4.8 | 464 | 16 Q8DWD1 | Q8DWD1 streptococ |
| 65 | 125.5 | 4.8 | 465 | 16 Q8ZP36 | Q8ZP36 salmonella |
| 66 | 125 | 4.8 | 462 | 16 Q8PHU0 | Q8PHU0 escherichia |
| 67 | 125 | 4.8 | 587 | 11 Q35055 | Q35055 rattus norv |
| 68 | 125 | 4.8 | 685 | 16 Q8EJ17 | Q8EJ17 shewanella |
| 69 | 125 | 4.8 | 1185 | 16 Q8EGS8 | Q8EGS8 shewanella |
| 70 | 124 | 4.7 | 391 | 16 Q9ZCM6 | Q9ZCM6 rickettsia |
| 71 | 124 | 4.7 | 501 | 8 Q94PS6 | Q94PS6 cryptocodi |
| 72 | 124 | 4.7 | 518 | 17 Q8THB8 | Q8THB8 methanosaic |
| 73 | 124 | 4.7 | 729 | 16 Q8F0Z0 | Q8F0Z0 leptospira |
| 74 | 123.5 | 4.7 | 445 | 16 Q8N9D5 | Q8N9D5 staphylococ |
| 75 | 123.5 | 4.7 | 432 | 16 Q8RGF7 | Q8RGF7 fusobacteri |
| 76 | 123.5 | 4.7 | 521 | 8 Q9TC6 | Q9TC6 nephroselm |
| 77 | 123.5 | 4.7 | 531 | 8 Q21293 | Q21293 reclinomona |
| 78 | 123 | 4.7 | 466 | 17 Q8TUI1 | Q8TUI1 methanosaic |
| 79 | 122.5 | 4.7 | 393 | 2 Q8KWT2 | Q8KWT2 bacillus su |
| 80 | 122.5 | 4.7 | 402 | 16 Q8P7D9 | Q8P7D9 xanthomonas |
| 81 | 122.5 | 4.7 | 465 | 16 Q8Z7G0 | Q8Z7G0 salmonella |
| 82 | 122.5 | 4.7 | 482 | 10 Q9SWR7 | Q9SWR7 cylindrothe |
| 83 | 122.5 | 4.7 | 512 | 16 Q9HVA7 | Q9HVA7 pseudomonas |
| 84 | 122.5 | 4.7 | 713 | 2 Q918D2 | Q918D2 polyangium |
| 85 | 122.5 | 4.7 | 1139 | 16 Q8ZC91 | Q8ZC91 yersinia pe |
| 86 | 122 | 4.6 | 643 | 17 Q29273 | Q29273 archaeglob |
| 87 | 121.5 | 4.6 | 482 | 10 Q9SWR6 | Q9SWR6 cylindrothe |
| 88 | 121.5 | 4.6 | 502 | 16 Q9CMU0 | Q9CMU0 pasteuralla |
| 89 | 121.5 | 4.6 | 534 | 2 Q9R6U5 | Q9R6U5 synechococc |

90 121.5 4.6 539 2 Q59631
 91 121 4.6 425 16 Q8RCV2
 92 121 4.6 523 16 Q8DAG1
 93 121 4.6 527 8 Q47563
 94 120.5 4.6 393 2 Q8KWS7
 95 120.5 4.6 467 8 Q99295
 96 120.5 4.6 472 16 Q990U7
 97 120.5 4.6 472 16 Q8NUG9
 98 120.5 4.6 482 10 Q8SMR8
 99 120.5 4.6 507 16 Q8XNM3
 100 120.5 4.6 512 10 Q94EC3

ALIGNMENTS

RESULT 1

Q9S6V3 PRELIMINARY; PRT; 528 AA.
 ID Q9S6V3
 AC Q9S6V3
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2003 (TrEMBLrel. 23, Last annotation update)
 DE Nucleoside triphosphate transport protein 1.
 GN NP1.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxId=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99138740; PubMed=9973346;
 RA Tjaden J, Winkler H.H., Schwoepe C., van der Laan M., Moehlmann T.,
 Neuhaus E.;
 RT "Two nucleotide transport proteins in Chlamydia trachomatis: One for
 net nucleoside triphosphate uptake and the other for the transport of
 energy.";
 RL J. Bacteriol. 181:1196-1202 (1999).
 DR EMBL, AJ010586; CAB39534.1; -;
 DR InterPro; IPR004667; ADP_ATP_car.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 SQ SEQUENCE 528 AA; 58039 MW; 8FBCDD5FC80B0E3 CRC64;

Query Match 81.5%; Score 2144; DB 2; Length 528;
 Best Local Similarity 77.9%; Pred. No. 7.7e-149;
 Matches 409; Conservative 53; Mismatches 49; Indels 14; Gaps 3;

QY 1 MKTEKEPKGKRSFLMPHITHEIKKVLPMFLMFCITFNVTYVLTDTKDTLLVGA PGSGA 60
 DB 1 MTQTEKPEPKKRSFLMPHITHEIKKVLPMFLMFCISFNVTLLDTKDTLLVTPGSGA 60
 QY 61 EAIPIKFWLVVPCAIIFMLIYAKLSNLSKQALFYAVAGTPELFFALFPVITYPLRDVL 120
 DB 61 EAIPIKFWLVVPCAIIFMLIYAKLSNLSKQALFYAVAGTPELFFALFPVITYPLRDVL 120
 QY 121 HPTPADRLQALIPGGLGLVAILKRWTPAAFYVLAELMGSVMLSMFGFANETIKHE 180
 DB 121 HPTPADRLQALIPGGLGLVAILKRWTPAAFYVLAELMGSVMLSMFGFANETIKHE 180
 QY 121 HPTPADRLQALIPGGLGLVAILKRWTPAAFYVLAELMGSVMLSMFGFANETIKHE 180
 DB 121 HPTPADRLQALIPGGLGLVAILKRWTPAAFYVLAELMGSVMLSMFGFANETIKHE 180
 QY 181 AKRFALFGIGANISLISGRAIIVASKLRASVSGVDPMGSLSLMLAMTIVSGLVMA 240
 DB 181 AKRFALFGIGANISLISGRAIIVASKLRASVSGVDPMGSLSLMLAMTIVSGLVMA 240
 QY 241 SYWVINKVNLTPRRYNEPEMOKGKGAAPKANNKDSFLYLDSPYILLTLVLAAYGIC 300
 DB 241 CYWVNNRVLTPRRYNEPEMOKGKGAAPKANNKDSFLYLDSPYILLTLVLAAYGIC 300
 QY 301 INLIVTWKSQLKQYPPNNNDYSEPMGNFSFWTGVSVILIMLVFGANVIRKGMTLGAIV 360
 DB 299 INLIVTWKSQLKQYPPNNNDYSEPMGNFSFWTGVSVILIMLVFGANVIRKGMTLGAIV 360

RESULT 2

Q9AS91 PRELIMINARY; PRT; 644 AA.
 ID Q9AS91
 AC Q9AS91
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Putative plastidic ATP/ADP-transporter.
 GN P0707D10.36 OR P0038D11.15.
 OS Oryza sativa (Rice), and
 OS Oryza sativa (Japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhacroidae; Oryzae; Oryza.
 OX NCBI_TaxId=4530, 39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=O. sativa; STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 clone: P0707D10.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=O. sativa (Japonica cultivar-group); STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 clone: P0038D11.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP002910; BAB40979.1; -;
 DR EMBL; AP003234; BAC05539.1; -;
 DR Gramene; Q9AS91; -;
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 SQ SEQUENCE 644 AA; 68960 MW; B3531863E48334AC CRC64;

Query Match 51.7%; Score 1358.5; DB 10; Length 644;
 Best Local Similarity 53.8%; Pred. No. 2.9e-91;
 Matches 275; Conservative 90; Mismatches 135; Indels 11; Gaps 8;

QY 7 KPFGLRSFLMPHITHEIKKVLPMFLMFCITFNVTYVLTDTKDTLLVGA PGSGA EAIPI 66
 DB 105 QPEKGK-KFL-GVEKTLKTIPLGIMFECILENTYITLDRFDVLVYAKSGSAEIIPEL 161
 QY 67 KFWLVVPCAIIFMLIYAKLSNLSKQALFYAVAGTPELFFALFPVITYPLRDVLHPTBEA 126
 DB 162 KTWVNLPMALIGFMLLYTLNLSREALFYVIFPELFFGAFVLPRLRVNHPPTALA 221
 QY 127 DRLQALIPGGLGLVAILKRWTPAAFYVLAELMGSVMLSMFGFANETIKHEAKRFYA 186
 DB 222 DKLALALPSPFLGPALIRIWSFCIFYMAELMGSVSVLFWGANOITTYEAKERYP 281
 QY 187 LFGLGANISLISGRAIIVASKLRASVSGVDPMGSLSLMLAMTIVSGLVMA SYWVINK 246
 DB 282 LFGLGANIALIFSGLTAVYFENIKRTLGPDGMEVSLKGMMSLVLLGLVITSYWGYN 341
 QY 247 KNLVLTDPFYNPEMOKGKGAAPKANNKDSFLYLDSPYILLTLVLAAYGICINLLEV 306
 DB 342 KVLINDPBL--PKSPRRKKKD-KPKLGKKEISLKVLSRYYVDLAVVAAIGISINLVEV 398

```

Qy 307 TWKSOLKL-OYNNNDYSEFMGNFSFWTGVVSLIMLFGVGNVIRKFGMLTGAULTVTPMV 365
Db 399 TWKSOLKL-OYNNNDYSEFMGNFSFWTGVVSLIMLFGVGNVIRKFGMLTGAULTVTPMV 457
Qy 366 LITGIVFPAVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKXALPDSTKEMAYI 425
Db 458 LITGIVFPAVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKXALPDSTKEMAYI 517
Qy 426 PLDDEOK-VKGAIDVVAARFGSGGALLIOGGLVITGSGTGMTPYLAIVLFLIAIWL 484
Db 518 PLDDEOK-VKGAIDVVAARFGSGGALLIOGGLVITGSGTGMTPYLAIVLFLIAIWL 577
Qy 485 VSATKLNLF--LAQSALKEQVADSDPA 513
Db 578 GAASSLDDKQSSSLAKEDLKRDMSAKKVDPS 608

RESULT 3
Q8MH67 PRELIMINARY; PRT; 671 AA.
AC Q8MH67;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Plastidic ATP/ADP transporter (Fragment).
GN AATPI.
OS Gaidieria sulphuraria.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC Gaidieria.
NCBI_TaxID=130081;
RN [1]
RP SEQUENCE FROM N.A.
RA Stemme C., Neuhaus E.H.;
RT "Analysis of the plastidic ATP/ADP transporter from the red algae
RT Gaidieria sulphuraria."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ251356; CAC80882.1;
DR InterPro; IPR004667; ADP_ATP_car.
DR Pfam; PF03219; TUC; 1.
DR TIGRFAMs; TIGR00769; AAA; 1.
DR Chloroplast.
KW NON TER
SQ SEQUENCE 671 AA; 73829 MW; 2025A868F2CFFFD7 CRC64;

Query Match 49.5%; Score 1302.5; DB 8; Length 671;
Best Local Similarity 53.2%; Pred. No. 3.7e-87;
Matches 259; Conservative 84; Mismatches 135; Indels 9; Gaps 5;

```

```

Qy 361 QASGLVAMFGTTPMLAVVGAIONILSKSTKXALPDSTKEMAYIPLDDEOKVKGKAID 440
Db 525 QASGLVAMFGTTPMLAVVGAIONILSKSTKXALPDSTKEMAYIPLDDEOKVKGKAID 584
Qy 441 VVAARFGSGGALLIOGGLVITGSGTGMTPYLAIVLFLIAIWLVSATKLNLF--LAQSAL 500
Db 585 VIGNPMKSGSGSFLIOGGLVITGSGTGMTPYLAIVLFLIAIWLVSATKLNLF--LAQSAL 642
Qy 501 KEQEVAAQ 507
Db 643 KRSEVAAE 649

RESULT 4
Q8MH67 PRELIMINARY; PRT; 588 AA.
AC Q8MH67;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Plastidic ATP/ADP transporter (Fragment).
GN Citrus hybrid cultivar.
OS Citrus.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Eucosids II; Sapindales; Rutaceae; Citrus.
NCBI_TaxID=171250;
RN [1]
RP SEQUENCE FROM N.A.
RA Li C., Weiss D., Goldschmidt E.B.;
RT "Citrus mRNA for plastidic ATP/ADP transporter."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY098893; AAM29152.1;
DR InterPro; IPR004667; ADP_ATP_car.
DR Pfam; PF03219; TUC; 1.
DR TIGRFAMs; TIGR00769; AAA; 1.
DR Chloroplast.
KW NON TER
SQ SEQUENCE 588 AA; 64735 MW; B4B6575BB36B37B CRC64;

Query Match 49.5%; Score 1302; DB 8; Length 588;
Best Local Similarity 51.1%; Pred. No. 3.5e-87;
Matches 261; Conservative 87; Mismatches 141; Indels 22; Gaps 5;

```

QY 424 YIPDQKVKGAIDVVAARFGSGGALIQOGLVIGSGIGAMTPYLAIVLFIITIM 483
DB 475 YIPDQKVKGAIDVVAARFGSGGALIQOGLVIGSGIGAMTPYLAIVLFIITIM 534
QY 484 LVSATKLNKFLAOGALKEQVADSDAPAS 514
DB 535 LGAARSLDTQF---TALROEELKEKEMERAA 562

RESULT 5

Q92J16 PRELIMINARY; PRT; 498 AA.
ID 092J16; 092J16;
AC 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE ADP carrier protein.
GN TICI OR RC0081.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Remesto-Audifiren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
RL Science 293:2093-2098(2001).
DR EMBL: AE008575; AAL02619.1;
DR InterPro; IPR004667; ADP_ATP_car.
DR Pfam; PF03219; TUC; 1.
DR TIGRFAMs; TIGR00769; AAA; 1.
KW Complete proteome.
SQ SEQUENCE 498 AA; 56661 MW; 975EB9C5BCC56B CRC64;

Query Match 40.9%; Score 1076; DB 16; Length 498;

Best local similarity 43.6%; Pred. No. 1e-70; Indels 20; Gaps 7;

Matches 213; Conservative 103; Mismatches 153; Indels 20; Gaps 7;

QY 11 KLRSLFMTHTHEKLVLPMLFPCITFNVTYVLRDTKTLIVGAPSGAEIPIKFWL 70
DB 12 ELKSIWIEREYENKFLPMAFMFCILNYSITLSIDGPFV--TDIGAEISLTKYI 69
QY 71 VVPCAIIFMLIYAKLSNLSKQALFYAVGTPFLIFALPFIYIPYLRDLHPTEFADRLQ 130
DB 70 VLPQVIMVIVYKLCIDILKQENVYVITSFPLGYFALFAFVLYPDLVHDPETISW 129
QY 131 AILPGLGLVAILRNWTFPAFYVLAELMGVSLMFWGPNATETTKHEAKRFALGCI 190
DB 130 SVAAPNVWTFRIYKQWSPASFYTAELMGVSLMFWGPNATETTKHEAKRFALGCI 189
QY 191 GANISILASGRAIYVASKLRASVSEGVDPWGISLRL---LMAMTIVGLVIMASYIN 246
DB 190 LANLALPTYSITIGCLHEKTOI-----VAHLKKEVPLFVIMTISPLVIILTRYMN 241
QY 247 KNVLTDPFYPNEEMOKGKGAIPRONKDSFLYIDRSPYILLTLTLVIAVICINLIEV 306
DB 242 KNVLTDPFYPNEEMOKGKGAIPRONKDSFLYIDRSPYILLTLTLVIAVICINLIEV 299
QY 307 TWKSGQLQYENMNDYSEFMGNFSFWGVSVLIMLFGGNAVIRKFGMLTGALVTPNVVL 366
DB 300 VMSKSVKLYLPTEKAYTITMGKFOFYQGVNVAIAFWL-IGSNILRVSWLTAAMITPLML 358
QY 367 LTGIVFFALVIFRNQ-ASGLVAMFGTTPMLAVVGAIONILSKSTKALPDSTKEMAYI 425
DB 359 ITGAFFAFIFPDSVIAHNLGILASGPLALAVMGIONVTSKVKVSLSDATGMAYI 418
QY 426 PLDQKVKGAIDVVAARFGSGGALIQOGLVIGSGIGAM--TPYLAIVLFIITIM 483
DB 419 PLDQKVKGAIDVVAARFGSGGALIQOGLVIGSGIGAM--TPYLAIVLFIITIM 478

QY 484 LVSATKLNK 492
DB 479 IYAVKGLNK 487

RESULT 6

Q8KPN2 PRELIMINARY; PRT; 514 AA.
ID Q8KPN2; 08KPN2;
AC Q8KPN2; 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Non-mitochondrial nucleotide transport protein.
GN NMT.
OS Holospora obtusa.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Holosporaceae; Holospora.
OX NCBI_TaxID=49893;
RN [1]
RP SEQUENCE FROM N.A.
RA Linka N., Hurka H., Lang F.B., Burger G., Winkler H.H., Stamme C.,
RA Urbany C., Seil I., Kusch J., Neuhaus E.H.;
RT "Phylogenetic relationships of non-mitochondrial nucleotide transport
proteins in bacteria and eukaryotes";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY120885; AAM0566.1;
DR InterPro; IPR004667; ADP_ATP_car.
DR Pfam; PF03219; TUC; 1.
DR TIGRFAMs; TIGR00769; AAA; 1.
SQ SEQUENCE 514 AA; 57531 MW; BEC3E24B9289E82F CRC64;

Query Match 38.8%; Score 1020; DB 2; Length 514;

Best local similarity 39.0%; Pred. No. 1.3e-66; Indels 36; Gaps 12;

Matches 206; Conservative 118; Mismatches 168; Indels 36; Gaps 12;

QY 1 MTKEERKPGKLRSGFLPDIHTEKLVLPMLFPCITFNVTYVLRDTKTLIVGAPSGA 60
DB 1 MSNLSRMSKWRRLWNPVPEEMSTFLPMVIVFSLFNTVARNVDSLVNPSGSG 60
QY 61 EAIPIKFWLVVPCAIIFMLIYAKLSNLSKQALFYAVGTPFLIFALPFIYIPYLRDL 120
DB 61 KVLPMVIVGALTTPCSIFAVIGYAKLSNLSKQALFYAVGTPFLIFALPFIYIPYLRDL 120
QY 121 H-PHEFADRLQAILPGLGLVAILRNWTFPAFYVLAELMGVSLMFWGPNATETTKH 179
DB 121 NLSVDWIKKQSIYPL-LKDFPVAAYWNYSLEFYMALMGVSLMFWGPNATETTKH 179
QY 180 EAKRFVALFGIGANISILASGRAIYVASKL--RASVSEGVDPWGISLRLMAMTIVGLV 237
DB 180 QAKRFYVYGWSNLGIVAAAGSLTAFSENFLEKPERLASGEKDFQRLRYCSCNVVACIA 239
QY 238 LMASYMINKNVLTDPFYPNEEMOKG--KKG--AKPRONKDSFLYIDRSPYILLTL 293
DB 240 IVSVYMWLNAKVLND-----AQKGTDAKGRSKKPLTLKESAVVYLTKRMYIAIL 291
QY 294 VIAIGICINLIEVWQSKLQYF-----NMNDYSEFMGNFSFWGVSVLIMLFGG 346
DB 292 MLAGGITINIVEVWQSKVGHYFSDPYTGLRDMNAVNAFPGKMFIGITITWVTLF-SK 350
QY 347 NVIRKFGMLTGALTTPVNVLLTGIVFFALVIFRNQ-ASGLVAMFGTTPMLAVVGAIONI 406
DB 351 NFGVSLGRPSANITTPVTMTIGGFLFVFKOYLGGLGCLFSTISLTVAGGLGVNV 410
QY 407 LSKSTKALPDSTKEMAYIPLDQKVKGAIDVVAARFGSGGALIQOGLVIGSGIGAM 462
DB 411 LTKGVKALFEPPTKEMAYIPLDQKVKGAIDVVAARFGSGGALIQOGLVIGSGIGAM 466
QY 463 GSIG---AMTPYLAIVLFIITIMVLSATKLNKFLAOGALKEQVADSDAPAS 507
DB 467 GGLGIPOLSVPLLGLGLVIGSGIGAMVAVNKL-ATAVAKVLETVQKE 513

RESULT 7
ID Q92HP9 PRELIMINARY; PRT; 501 AA.
AC Q92HP9; 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ADP carrier protein.
OS Trichomonas vaginalis.
OC Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxId=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RL "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
RL Science 293:2093-2098 (2001).
DR EMBL; AE008630; AL03260.1; -
DR InterPro; IPR004667; ADP_ATP_car.
DR Pfam; PF03219; T1C; 1.
DR TrEMBL; TIGR00769; AAA; 1.
KW Complete proteome.
SQ SEQUENCE 501 AA; 57104 MW; 2EC589FAD865EA3B CRC64;

Query Match 37.1%; Score 976.5; DB 16; Length 501;
Best Local Similarity 40.0%; Pred. No. 1.9e-63;
Matches 199; Conservative 106; Mismatches 174; Indels 19; Gaps 8;

QY 9 FGKLRSPFPIHTEHLKVLPMFELFCITFNTVLDRTKTLVGAAGSGAEAIPIKF 68
DB 8 FEKVEIMWPIERKELKFIPLMLMLCILFNFGLRSIKDSLIV--PSMGAEIISFLKL 65
QY WLWVPCALIFMLIYAKLSNLSKQALFYAVGTRPILFALPFTVYLPRLDVLPHT-EPAD 127
DB 66 WLWVPCALIFMLIYAKLSNLSKQALFYAVGTRPILFALPFTVYLPRLDVLPHT-EPAD 125
QY 128 RLQALPGLGLVLAIRNMTFAFYVLAELMGVSLMGWGFANETIKHEARKFYAL 187
DB 126 KLISY-ENFKMFIRIGSOWSALMYIFAEILMSAVINLMFQGFANHLFDISKARFIPV 184
QY 188 FGIGANISLLASGRAYWASKLRASVSGVD-----WGISRLIMAMTVSGVLV 239
DB 185 LGMVGNIGLIIAGSVLVFFSSGQDIVIDSELLPDSFNSAGNAIMLQPIMSIIVTAGI 244
QY 240 ASYMNKNNVLDPRFVNEEMQKKGAKPRMNMKDSFVLVDRSPYLLLTLLVIAIGI 299
DB 245 LLEFRINFIPLDS--INVLDKAKVTAAKTKTSLVIESIKVHSHKYGRIALLLICGL 302
QY 300 CINIIEVTWKSQKLQYENNDYSEFMGNFSFMTGVSVLIMLFVGNVIRKFGWLTGAL 359
DB 303 LNIIVEGWKAKIKELHNTIDYVFMGRFNIMWGISCVTMI-IGSNILRLGLISAL 361
QY 360 VTPVWVLTGIVFPALVIFRNOASGLVAMFGTTPMLAVVGAIONISKSTKALPST 419
DB 362 LTPIMLSITGMFMFIIFIEBIGECFDPNL--LYAAIIVGAIQNIISKSKSYLFPST 419
QY 420 KEMAYIPLDOOKYKGAIDVVAARPGSGALIQGLLVY--CGSIGAMTPYAVLL 477
DB 420 KEMAYIPLDOOKYKGAIDVVAARPGSGALIQGLLVY--CGSIGAMTPYAVLL 479
QY 478 FIATMLVSAATKLNKLF 495
DB 480 VAMSLIMINVIKLNKEYV 497

RESULT 8
ID Q9X9D2 PRELIMINARY; PRT; 540 AA.

AC Q9X9D2; 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Nucleoside triphosphate transport protein 2.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydia.
OX NCBI_TaxId=813;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99138740; PubMed=9973346;
RA Tjaden J., Winkler H.H., Schweeppe C., van der Laan M., Moehlmann T.,
RA Neuhaus E.;
RL "Two nucleoside transport proteins in Chlamydia trachomatis: One for
RT net nucleoside triphosphate uptake and the other for the transport of
RT energy";
RT J. Bacteriol. 181:1196-1202 (1999).
DR EMBL; AJ010587; CAB39535.1; -
DR InterPro; IPR004667; ADP_ATP_car.
DR Pfam; PF03219; T1C; 1.
DR TrEMBL; TIGR00769; AAA; 1.
SQ SEQUENCE 540 AA; 59648 MW; C43C2A3FCA4226FD CRC64;

Query Match 36.5%; Score 960.5; DB 2; Length 540;
Best Local Similarity 37.5%; Pred. No. 3.1e-62;
Matches 198; Conservative 115; Mismatches 194; Indels 21; Gaps 6;

QY 4 TEKRPFGKLRSPFPIHTEHLKVLPMFELFCITFNTVLDRTKTLVGAAGSGAEAI 63
DB 3 SEVSPFSKFRYPPIRSEFSKPIPLFLAFVGVNVALKTRDSLIVGSRGAIEVI 62
QY 64 PFIKWLVPVCAIIFMLIYAKLSNLSKQALFYAVGTRPILFALPFTVYLPRLDVLPHT 123
DB 63 PFLWVGIVPQAVIVTMIYGMMSRRYSRGVIFSLVGGFLGFALFVAVIPIGPAHLN 122
QY 124 EFADRLQALPGLGLVLAIRNMTFAFYVLAELMGVSLMGWGFANETIKHEARK 183
DB 123 KLAKLOSILPDRGGRGVWQVWSYSLIYVMSLSVSLTLPFGVANHITSVREAGR 182
QY 184 FYALFGIGANISLLASGRAYWASKLRASVSGVD-----WGISRLIMAMTVSGVLV 242
DB 183 FYALINGLMSSVFAVGSVSLMGRSPVIAFPAVDPHEMLNITLLIIVLAGGVITLVY 242
QY 243 WINKRVLTDRFPNPE-----EMQKKGAKPRMNMKDSFVLVDRSPYLLLTLL 293
DB 243 QKLR--LMDSTMLIEGLAEMSVADLKKKKRSKAK--AKSFLALIRSRYLIGIAVV 298
QY 294 VIAIGICINLEVTWKSQKLQYENNDYSEFMGNFSFMTGVSVLIMLFVGNVIRKFG 353
DB 299 VLSNVLVHLEEVWVKQVCRIVASRVFNSYMSRITTLTGIVSALGIFPAAGOTIRWG 358
QY 354 WLGTALVTPVWVLTGIVFP-ALVIFRNOASGLVAMFGTTPMLAVVGAIONISKSTK 412
DB 359 MTGVALVPLRITLLITGALFEGAIYAVGDMNIFGIIIGISPLVITAMLGQVQNFSAIK 416
QY 413 YALDSTKENAYIPLDOOKYKGAIDVVAARPGSGALIQGLLVYCGSIGAMPYL 472
DB 419 FTYPDQTEKMAFIPLDEDEKRYGKAIDGVISRVKSGGSIVYQGLLIIFSSVAASINAI 478
QY 473 AVILLFIATMLVSAATKLNKLF 495
DB 479 TIVLLLAGSWIFVIAMGRIYAKTETLVVNASBEDVLQERBAS 526

RESULT 9
ID Q9ZG15 PRELIMINARY; PRT; 499 AA.
AC Q9ZG15; 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ADP carrier protein.

GN TLCS OR RC1138.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii."; Science 293:2093-2098 (2001).
 RL EMBL; AE008663; AA03676.1; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; T1C; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KM Complete proteome.
 SQ SEQUENCE 499 AA; 56918 MW; CD873E469763E2AF CRC64;
 Query Match 34.5%; Score 907.5; DB 16; Length 499;
 Best Local Similarity 37.4%; Pred. No. 2.2e-58;
 Matches 190; Conservative 107; Mismatches 176; Indels 35; Gaps 13;
 QY 1 MKTKEKRP-GKLRSLPMTIHTELKKVLPMFLMPCITENYTVLRDTKDTLIVGAPSG 59
 DB 1 MISTSRSRKPKFRAFPVHVEYLGKFTPKSTLFCILFNQNVRLKLSISE--IS 58
 QY 60 ABAIFIKFWLVVPCALIFMLIYAKLSNLSKQALFYAVGTPELLIFPALFPYIYPLRDV 119
 DB 59 ABIAFPKYYCVTPAALFVLIYAKMINVLEFEKIFYLSAFISFPVFTFYIYPNHI 118
 QY 120 --LHTEPDRLOALIPGLGLVALNRWTPAAYVLAELMGVSLMFGFANEITK 177
 DB 119 FVHHNNLADWERY--PHFKYISLVGNGYIVYSIAELWPNIFYVLLFVQFANELIT 176
 QY 178 IHEARFPYALPGIGANISLSASRAIVAS-----KLRAVSEG-VDPWGISLTLAM 230
 DB 177 TEEARKFTYFLSPGNSLLIVGFLMMNLISSEDTIIFKFMSTISDSKITLVQSTIVAI 236
 QY 231 TIVSGLVMAASYWYINKVLTDPFRYPNPEMOKGKG--AKPKMMKDSFLYLDSPYL 288
 DB 237 AITCLVLR---FISKVFTNPLFY-----AKASGRSTSERMGLISFKIAKSKLM 287
 QY 289 LITLIVAVGICINILEVTWKSQKLQYPPNNDYSEPMGNFSPFTGVVSVLIMFVGAV 348
 DB 288 LLLISAAFGPAINLVEAVWKAKEIYPTVNTVAEFNSLYILMTG-VAIMVMTIIGNNI 346
 QY 349 IRKFGMLGALVTPVWVLTGIVFPALVIFRQA-----SGLVAMGTTPMLAAVVGAIQ 404
 DB 347 MRMHMPPAAVAVSPIYIMWTGILFVLIYFDQIISLPGAILM--SPALAVSIGIQ 403
 QY 405 NILSKSTKALFDSTKEMAYIPLDOEKVKGKAIDVVAARFGKSGALLIQGL--LVYC 462
 DB 404 NILAKTKYSIMWTSREMLYIPLDEELKTKGKAADVISAKEYSSGSLVQSIITLVPT 463
 QY 463 GSIGAMTPYLAIVLFIATIMLVSAATKL 490
 DB 464 ATFTLISPLMVVFTFVCLAMIVAVRKI 491
 RESULT 10
 Q92HV4 PRELIMINARY; PRT; 511 AA.
 AC Q92HV4;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE ADP_ATP carrier protein.
 GN TLCS OR RC0666.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii."; Science 293:2093-2098 (2001).
 RL EMBL; AE008626; AA03204.1; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; T1C; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KM Complete proteome.
 SQ SEQUENCE 511 AA; 58036 MW; E194D7E21591EB02 CRC64;
 Query Match 34.5%; Score 907.5; DB 16; Length 511;
 Best Local Similarity 37.9%; Pred. No. 2.2e-58;
 Matches 187; Conservative 111; Mismatches 175; Indels 21; Gaps 9;
 QY 9 FGKRSFLMPPIHTELKKVLPMFLMPCITENYTVLRDTKDTLIVGAPGAAIPIKF 68
 DB 19 FSKLTDYIPIKRHSKDFLTITLMPFCLIFQNLIRALKDSIVTM--IGAEIISFKF 76
 QY 69 WLVPICALIFMLIYAKLSNLSKQALFYAVGTPELLIFPALFPYIYPLRDVH--PTEFA 126
 DB 77 WGVMSAPLMTAIYKLVKKNKKAENIFYLITSLTFPFAAYVIFPHENLHSPVTQ 136
 QY 127 DRLQALIPGLGLVALIRNMTFAAFYVLAELMGVSLMFGFANEITKHEAKRFA 186
 DB 137 NLMAST--PNLKWPIWLSKSPSLFYIALMNNVYVALLFQFVNNITVESKRYTP 194
 QY 187 LFGIGANISLSASRAIVASKLRASVSEGVDPNG---ISRLIMANTIVSGLVASY 242
 DB 195 LFGILSGTGIYALQCFLENLSINDYVTKFKALOSSFHTLSIQIITLVLIIGIATKF 254
 QY 243 WINKNVITDPRFRPNPEMOKG--KGAKPKMMKDSFLYLDSPYLITLIVAVGIC 301
 DB 255 WLNHRKVL-----DKEMLALRFKAKKSKWTIASEFQMLSSHRIATLTCYGIAI 308
 QY 302 NLEVTWKSQKLQYPPNNDYSEPMGNFSPFTGVVSVLIMFVGAVIRKFGMLTALVT 361
 DB 309 NLVSGPMKAATKTKYKPTFAAIFGSLYSTGVFTILFV--LSNIVRRLGWTAAVIT 367
 QY 362 PNVVLTGIVFALVIFRQNSGLVAMP-GTTPMLAVVGAIONLSKSTKVALFDSTK 420
 DB 368 PLIVFTIGILFFAVNNERFAGLIIANFILDPAIIATTGAIQNVLSKSKYLPDSTK 427
 QY 421 EMAYIPLDOEKVKGKAIDVVAARFGKSGALLIQGLLVCGSIGAMTPYLAIVLFI 480
 DB 428 EMAYVPLDPETIKIGKAADYIGTKLGKSGAFQSLVFIILPSASIOSISTCMLITFI 487
 QY 481 A--IMLVSAATKLANK 492
 DB 488 TCLTMATKALANK 501
 RESULT 11
 Q92198 PRELIMINARY; PRT; 507 AA.
 AC Q92198;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE ADP_ATP carrier protein.
 GN TLCS OR RC0522.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renseto-Audifren P., Fournier P.-E., Barbe V.,
 SA Sauton D., Roux V., Coesart P., Weissenbach J., Claverie J.-M.,
 RA Raulot D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
 RL Science 293:2093-2098(2001).
 DR EMBL; AB008614; AL030601; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; T1C; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KW Complete proteome.
 SQ SEQUENCE 507 AA; 58365 MW; 49A92CD91E300C70 CRC64;

Query Match 31.4%; Score 827; DB 16; Length 507;
 Best Local Similarity 35.8%; Pred. No. 1.7e-52;
 Matches 179; Conservative 106; Mismatches 183; Indels 32; Gaps 9;

QY 11 KLRSEFLMFIHHEKLVPMFLMFCITENTVTRDTDLVGAAGSAGAIPIKFWL 70
 DB 18 KFRRIWPIRISYELTKFIPMLLMFILLNOLVRSTKDSFVTLIS--EVLSPKLMG 75
 QY 71 VVPCAIIFMLIYAKLSNLSKQALFYAVGPELFFALFPTVIYPLRDVHPTEFADRLQ 130
 DB 76 EMPGILFVIFYSKLONMTTEQVFRITGTFFFAIFGFLIPYRHFPHDELIKH 135
 QY 131 AILPGLGLVAILRNTPFAFVYLAELGSMLSMFPGANETIKHEAKRYALFGI 190
 DB 136 ITVPLHLMFLINGQMSLVLFYINGELMPVIFLLWOLANKITKVEAPRYSFPTL 195
 QY 191 GANISLASGRAIYAWASK---LRASVSEGVDPWGISRLMAMTIVSGILMASYWIN 246
 DB 196 FQGNLNLISGVITFYFAKSEHFLPLSHLNDNEILKSTIYLLSGLCALHLKID 255
 QY 247 KNVLTDRFYNPBEWQKKGAKP---KNNKDSFLYIDRSFYILLTLVIAIGICI 301
 DB 256 KSVV-----EADKNIFKQNMIDILKLSVDASAVILTSRYLGICLWMSYSMSI 306
 QY 302 NLIEVTKSQKQYQPMNDYSEFMGNFSFTGVSVLIMLFVGNVIRKGMTALVT 361
 DB 307 SLIEGLMWSKVKQYLPATKDFIAHGVKVFMTGILT-LVSAFLGSSILRIGCMFWGAIIT 365
 QY 362 PVWVLGLGVFAVIFPNQASGLVAMFG-TTPMLAVVGAIONILSKSTKVALPDSK 420
 DB 366 PIMFGAGVWFSTVEENHGNIVNTIGAPLVIVIFGGLMHWLSKYSLSLPAIK 425
 QY 421 EMAYIPLDQOKVKAIDVVAARFGKSGALLQOGLLVICGSI--GAMTPYLAIVLLF 478
 DB 426 EMVYIPLDSEKTKGKAADVDMGAKIGKSIGAITQ---FISFIFPAIHNDIAGLLMF 481
 QY 479 ----TIAIWLVSATKLNLF 494
 DB 482 SFIIICLLMVGKVLISKY 501

RESULT 12
 Q8SUG0 PRELIMINARY; PRT; 543 AA.
 ID Q8SUG0
 AC Q8SUG0
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein ECU10_0520.
 GN ECU10_0520.
 OS Encephalitozoon cuniculi.
 OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
 OX NCBI_TaxId=6035;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RA Genoscope;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1.
 RX MEDLINE=21576510; PubMed=11719806;
 RA Katicka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
 RA Prenslér G., Barbe V., Peyreallade E., Brotier P., Wincker P.,
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
 RA Weissenbach J., Vivares C.P.;
 RT "Genome sequence and gene compaction of the eukaryote parasite
 RT Encephalitozoon cuniculi".
 RL Nature 414:450-453(2001).
 DR EMBL; AL590449; CAD25771.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 543 AA; 61647 MW; 44D7AB5F2A55B1A CRC64;

Query Match 13.9%; Score 366; DB 5; Length 543;
 Best Local Similarity 23.7%; Pred. No. 9.3e-19;
 Matches 125; Conservative 119; Mismatches 228; Indels 56; Gaps 15;

QY 5 EEKPFGLRSEPL-----WPIHTEKLVPMFLMFCITENTVTRDTKTLVIGAP 56
 DB 23 EEEVYKGTGPFKHIRVARNMP-----RVLYSLILFGVITWHTMGILREMYLMGR 74
 QY 57 GSGAEALPIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGPTFLIFPALFPTVIYPL 116
 DB 75 -QPMSEMFITKSTFLPFCISLIFTMAIQGLSLPSPMFTITLILFSGCYILFGLVWPL 133
 QY 117 -----RDVLPTEPAD-RLQAILPPGLGLVAILRNTPFAFVYLAELGSMVL 164
 DB 134 KGYIQKDFWYSRDI-----FGDGMESLRHLPLYPVLVFNEMTSSFLPCEMKGALV 188
 QY 165 SLMPWGANETIKHEAKRYALFGANISLASGAIAIYAWASKLRASVSEGVDPWGIST 224
 DB 189 SYEFNIFANVSRRQSORIVYNISNALSIFLSAVLTLVFNKMRGVAFETKELEFRI 248
 QY 225 RLIMAMTIVSGILMASYWINKNVLTDRFYNPBEWQKKGAKPNNKDSFLYIDRS 284
 DB 249 LILVLSGTVGILAKKY--MEREILPAPVFL-IREVEK-TSERRLKIDEARQTSRS 304
 QY 285 PYILLTLVIAIGICINLIEVTKSQKQYQPMNDYSEFMGNFSFTGVSVLIMLFV 344
 DB 305 KLIIAISLNLTLVGVSTLVEATFKSGIAGARTNNSKETPANFYNGLBQIIAISLV 364
 QY 345 GGN-----VIRKGMVLGALVTPVMVLTIGVFPALVIFNQA--SGLVAMFGI----- 391
 DB 365 VINTPVSALVKKGWKLIAL-PIVIAFSLFSVFLAFYVAGADSGANVLPGLFQRM 423
 QY 392 TPIMLAVVGAIONILSKSTKVALPDSKEMAYIPLDQOKVKAIDVVAARFGKSGG 451
 DB 424 PTILENTLGLVTRNASKIKIKYGADVSKAISQIDPLYRAKYKAVYDGLCGKLGSLG 483
 QY 452 ALI---QOGLLVICGSIAMTPYLAIVLLFIIAIIWLVSATKLNLF 496
 DB 484 SIICVTWGLMDIT-DIRAVSVSGILIVIIIMWVFLKYLRSQFA 530

RESULT 13
 Q8SUP9 PRELIMINARY; PRT; 553 AA.
 ID Q8SUP9
 AC Q8SUP9
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein ECU10_0540.
 GN ECU10_0540.
 OS Encephalitozoon cuniculi.
 OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
 OX NCBI_TaxId=6035;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RA Genoscope;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

[2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=GB-M1;
 RC MEDLINE=21576510; PubMed=11719806;
 RX Katiinka M.D., Duprat S., Cornilliot E., Metenier G., Thomarat F.,
 RA Prenslar G., Barbe V., Peyretailade E., Brotier P., Wincker P.,
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
 RA Weissenbach J., Vivares C.P.;
 RT "Genome sequence and gene compaction of the eukaryote parasite
 RT Encephalitozoon cuniculi."
 RL Nature 414:450-453(2001).
 DR EMBL; AL590449; CAD25773.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 553 AA; 62160 MW; 5091200D1F656BF5 CRC64;

Query Match 13.3%; Score 351; DB 5; Length 553;
 Best Local Similarity 25.0%; Pred. No. 1.3e-17;
 Matches 132; Conservative 89; Mismatches 236; Indels 70; Gaps 17;

QY 14 SFLMP-----ITHELKVLPMFLMFCITENYTVLRDTKDTLIVGAGSGAEALPFIKF 68
 Db 29 STWPLSRIRVRCCKWKLMSLAFIFGASAVIYFSRVMKDSFVLSR--QPLAISFLKT 86
 QY 69 WLVPICALIFMLIYAKL--SNLSKQALFYAVGTFFLIFALFPTVIYPLRDVLP-----122
 Db 87 CFVLPISIVTIGVOKLAVTRISKVPDTLIAFSFL--YLLIGNVLLPFAKTIQPGUYF 144
 QY 123 --TEPADRLQAILPPG--LLGLVALRNMTFAFYLAELMGSVMLSMFWMGPAHEITK 177
 Db 145 SRDIFADDMKAY--KGFEFLFALIFLNMWITSFYVCAELFGSLIVQFMFLAFANALT 202
 QY 178 IHEAKRFALFGIGANISILASGRAIIVASKLRASVSGVDPWGISLRLMAMTI--VSG 235
 Db 203 IROSTRMPLFYVSNILLSSSESTFSYK-----KVRWDYKTKCLTNSFEAFVG 255
 QY 236 LVLMAZYW--WINKNVLTDPREFYNEEMQKGAKEPKMMKDSFLYIDRSPYLLTL 292
 Db 256 AMIATVLYKRYAEETILKKQLFITEGVAK--KKGRKSAGSSEMKLMAQSKPLVANVM 314
 QY 293 LVIAVGICINLIEVTWKSQKLQYNNNDYSEPMGNSFWTGVVS-----VLIML 342
 Db 315 NALFYAGTNTLIESMKNGISVAADANN--MEKRAYSASIVSGEQRVVALVALITLL 369
 QY 343 FVGGNVIRKFGMLNATLPVWVLLTGIVFPALVIFR--NQASGLVANFGTTPMLAVV 400
 Db 370 TPISLTVQTHGMITWAIIVPLVTLVSLVIFGSAFENSNYPEG---KTSVILSSLVK 424
 QY 401 GAIQNIL-----SKTKYALFDPSTKEMAYIPLDOEKVKGKAIDVVAARFGK 448
 Db 425 GYKRFNIECNIGIYCVSGMKIARAFYDISKEALSLOIDPLRYRRLKAVYDGLCGKLGK 484
 QY 449 SGGALLIQGLLVI--CGSIGAMTPYLAVILLFIATLVSATKLNKF 494
 Db 485 SIGSLYAMFMSVGVNDVRAAAPTILGWMILISPIWISVILNRY 531

RESULT 14
 Q8SRA2 PRELIMINARY; PRT; 559 AA.

AC Q8SRA2;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE ADP/ATP carrier protein 1.
 GN EC008_1300.
 OS Encephalitozoon cuniculi.
 OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
 OX NCBI_TaxID=6035;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RA Genoscope;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

[2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=GB-M1;
 RC MEDLINE=21576510; PubMed=11719806;
 RX Katiinka M.D., Duprat S., Cornilliot E., Metenier G., Thomarat F.,
 RA Prenslar G., Barbe V., Peyretailade E., Brotier P., Wincker P.,
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
 RA Weissenbach J., Vivares C.P.;
 RT "Genome sequence and gene compaction of the eukaryote parasite
 RT Encephalitozoon cuniculi."
 RL Nature 414:450-453(2001).
 DR EMBL; AL590448; CAD26436.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 559 AA; 63516 MW; 9B1558ED28BCD5D CRC64;

Query Match 12.9%; Score 338.5; DB 5; Length 559;
 Best Local Similarity 24.6%; Pred. No. 1.1e-16;
 Matches 134; Conservative 107; Mismatches 237; Indels 67; Gaps 19;

QY 10 GKLSPLMPIHTEHLKVLPMFLMFCITENYTVLRDTKDTLIVGAGSGAEALPFIKF 69
 Db 31 GILRYF--RVARAAYTFALIGLMFGIIGFYSMRLIKDMFWN--VRQETTLFIKIF 86
 QY 70 LVVP--CAIIFMLIYAKLSNLSK-----QALFYAVGTFFLIFALFPTVIYPLRD 118
 Db 87 YLFPVSMALVPLIQYMLGTQVSRIFSLFCGGSFSLFLGAVFLIEHQVSPSK-FLPRD 145
 QY 119 VLAHTEPADRLQAILPPGGLGLVALRNMTFAFYLAELMGSVMLSMFWMGPAHEITK 178
 Db 146 MFIDGKSSRRLNFKSMFLNMEPLATYF--ISAEMWGSILVSLYFLNESCCTI 201
 QY 179 HEAKRFALFGIGANISILASGRAIIVASKLRASVSGVDPWGISLRLMAMTI--VSG 237
 Db 202 RQSRFIPPLIITNVSLFSAIYAGAFPLREKLA-----FQGNVLLSGIFFGELV 256
 QY 238 LMAZYW--WINKNVLTDPREFYNEEMQK--KKGAKEPKMMKDSFLYIDRSPYLLTL 293
 Db 257 VLVIKFLYIERVYMKRPLF---VSSGSRKKAANVSFSEGLIEMSGKLLAMSLI 312
 QY 294 VIAVGICINLIEVTWKSQKL--QYNNNDYSEPMGNSFWTGVVS-----VLIML 347
 Db 313 VLFNINISYNWSESTFKYGVKAAEYFN--EEKGKXSGFENRIDQWTSVVVLCNLSPPSS 371
 QY 348 VIRKFGMLTGVLPVWVLLTGIVFPALVIFR--NQASGLVANFGTTPMLAVV 400
 Db 372 YVETRGFLVGLIPIYTLWAIIVPLGSALYNTMESGGLIYVGLFPGKPLVLENYF 431
 QY 401 GAIQNILSKTKYALFDPSTKEMAYIPLDOEKVKGKAIDVVAARFGKSGGALLIQGLLV 460
 Db 432 GVIFMSLLKITKYSAFPICKKGLMRINPTVBARFKSVYDGIFFGLKSGISGI--YGLIM 489
 QY 461 I---CGSIGAMTPYLAVILLFIATLVSATKLNKF-----LAQSALKEQ 503
 Db 490 FEALDTEDLRKATPTTGAIIFFIVWVKAIILYSRYESAQVQHNRDVIDIMTEKARKSL 549
 QY 504 EVAQE 508
 Db 550 ETPEE 554

RESULT 15
 Q8SUG7 PRELIMINARY; PRT; 536 AA.

AC Q8SUG7;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein EC010_0420.
 GN EC010_0420.
 OS Encephalitozoon cuniculi.
 OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
 OX NCBI_TaxID=6035;
 RN [1]
 RP SEQUENCE FROM N.A.

SQ SEQUENCE 441 AA; 47931 MW; BDD996A58EC4E89 CRC64;
 Query Match 8.7%; Score 229.5; DB 16; Length 441;
 Best Local Similarity 23.1%; Pred. No. 7.7e-09;
 Matches 122; Conservative 84; Mismatches 197; Indels 125; Gaps 27;
 3 KTEEPFKLMSFLPIHTELKYLPMFLMFCITFNVTYLRDQKDLIVGAPSGSGBA 62
 Db 22 RPEEP-----AVLM-----SMLYVALFLAY-----VLRPRDEL--GVAG-GVQN 61
 Qy 63 IPFKFWLVPCALIFMLIVAKLSNLSKO-----ALFYAVGTPFLIFALPFTVIYP 115
 Db 62 LP-----WLFSTG-TLLAMLVASPLFALAVRSRPRQFIALAVRFPAANLVFLAL----- 109
 Qy 116 LRDVLPPEFADRLQALIPGLGLVALLRNWTFAFYVAELMGSVMSLMFGFANEI 175
 Db 110 ---LH---FADPMQV-----WVGRAFPIWVSFNLFFVVS-VFMSFWMDL 148
 Qy 176 TKIHEAKRFYALFGANISLLASGRATVWASKLRASVSEGV--PMGISLRLLAMTIV- 233
 Db 149 FDSEGGKRLFGIFAAAGT---AGS---LIGSAITSGIIEHIDRW-----LMAIATVF 195
 Qy 234 SGLVLMASVWYINKVLTDPREFY-----NPEMOKGKGAAPKKNMKSPLYLDRSPYI 287
 Db 196 LELAVLAS-----RLRSRIAPAFQHARNDPQPLG-----GIFAGVHTLSRPLY 243
 Qy 288 LLLTLVAVGICINLIEVTWKSQULKQYPMNDYSEMGNSFMTGVASVLMFVCGN 347
 Db 244 GGLAIFILLYSVTSTFLYFOQASIAQASFPDPAATAFANIDLVAIVTLVFOLFYVGR 303
 Qy 348 VIRKGMLTGALVTPVWVLTGIVFPAIVFRNAGSVAMEGTPMLAVVAGIQNIL 407
 Db 304 MVAIVGIVATCVLPVLSL-----GPAAL-AASP-SVAIVVAH--OVA 343
 Qy 408 SKSTKYALFDSTKEMAYIPLDQOKVGAKAIDVVAARFGSGALIGGLVTCGIGA 467
 Db 344 RRVANFALARPARELFTSSAREDRYKAKNFIDTVVGRGQDQIASWGVAIGML-GLTLA 402
 Qy 468 MTPPLAVILFPIALVWVSATKANKLFLAQSALEQEAQEDSAPASS 515
 Db 403 QIPMAIVPL---SAWIGLSVLTGRTHQAO-----RQDAATAPS 439
 RESULT 18
 Q9P6U3 PRELIMINARY; PRT; 454 AA.
 AC Q9P6U3;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Hypothetical protein KCC2872.
 GN KCC2872.
 OS Xanthomonas campestris (pv. campestris).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 NC NCB1_TaxID=340;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33913 / NCPBB 528;
 RX MEDLINE=20202145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 Alves L.M.C., do Amaral A.M., Bercolini M.C., Camargo L.B.A.,
 Camarotte G., Cannavan F., Cardoso J., Chambeiro F., Clapina L.P.,
 Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 Formighieri E.F., Franco M.C., Greggio C.C., Guber A.,
 Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 Martins E.C., Medeiros J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,

RA Trindade dos Santos M., Truffi D., Teal S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.,
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities.";
 RL Nature 417:459-463(2002).
 DR EMBL; AB012400; AAA42144.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 454 AA; 49344 MW; 405A41811E056D7 CRC64;
 Query Match 7.7%; Score 203.5; DB 16; Length 454;
 Best Local Similarity 19.8%; Pred. No. 6.3e-07;
 Matches 101; Conservative 78; Mismatches 152; Indels 179; Gaps 24;
 31 FLMFECTFNVTYLRDQKDLIVG-----PGS-----GAEATPFKPMVYV--PCALI 77
 Db 30 FLYFFCLSGYYVLRPVREMGASADVAAIFPAGMIAPFGAHGHPDLQTLQVFTCTFL 89
 Qy 78 FML-----YVAKLSNLSKQALF-----YAVGTPFLIFALPFTVIYPLRDVLPTEFADR 128
 Db 90 ITLLQPIYGALVSRVRRVFLPVYGFPIAT-LLLFYVFDIGV----- 133
 Qy 129 LQALPPGLIGVALIRNWTFAFYVAELMGSVMSLMFGFANEITKHEARFPALE 188
 Db 134 ---FG-RGMAFFLWTGVFNLFV-----AVFMSFVADVFSNAEARSYYGYI 175
 Qy 189 G-----IG-ANISLASG-----RAIV--ASKLRASVSE 215
 Db 176 GAAGTGAFLGPIITRLTVRIGIALHMLVSAGLGMCCVCLRLMAVAREEROVTS 235
 Qy 216 GVDPWGISLRLLAMTIVSGVLMASVWYINKVLTDPREFYENPEMOKGKGAAPKKNMK 275
 Db 236 GEV-----MGDVLTGKLTIV----- 252
 Qy 276 DSFLYDRSPYLLTLVAVGICINLIEVTWKSQULKQYPMNDYSEMGNSFMTGW 334
 Db 253 -----REPLRLWAMVW-FGVGVGTLLYNEQALVRLTYDAASTAYVANIIDLAVN 304
 Qy 335 VVSIVLMFVGVGNVIRKGMVLTGALVTPVWVLTGIVFPAIVFRNAGSVAMEGTPML 394
 Db 305 ALTVLQTLVLRALLSFFG-----IAPLLIFGAALIFGPVAVL--SASPL-----PM 349
 Qy 395 MNAVVAIONILSKSTKYALFDSTKEMAYIPLDQOKVGAKAIDVVAAR----- 445
 Db 350 MNAV-----QWTRRSSEPALAKPARETLTVRDREVRVYAGATIDVVVRRGGLSVVMI 404
 Qy 446 -----FGKSGGALIQGLLVTCG-STGAM 468
 Db 405 HKVVSAPGSS--VVFVGVLVACGMVVGAL 432
 RESULT 19
 Q9FLC3 PRELIMINARY; PRT; 496 AA.
 AC Q9FLC3;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE Genomic DNA, chromosome 5, TAC clone K18123.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCB1_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Colombia;
 RX MEDLINE=98344145; PubMed=9679202;
 RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asanizu E., Miyajima N.,
 RA Tabata S.,
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence
 RT features of the regions of 1,381,565 bp covered by twenty one
 RT physically assigned P1 and TAC clones.";
 RL DNA Res. 5:131-145(1998).

DR EMBL: AB010692; BAB09974.1; -
SQ SEQUENCE 496 AA; 53606 MW; 7BA868EE7369CDE6 CRC64;

Query Match 7.5%; Score 197; DB 10; Length 496;
Best Local Similarity 22.0%; Pred. No. 2.1e-06;

Matches 119; Conservative 78; Mismatches 203; Indels 140; Gaps 26;

```

QY 19 IHTHLLKVLPMPLFEFCITFNTVTLRDTKDLIVGAPSGAALP--PI-KFWLVPCA 75
DB 17 VHPHLLPALHSSSCFFFLSAFVVLPLRDE---GALSLSKSLPGLFVGSFLITLAA 73
QY 76 IIFMLYAKLSNLSKQAL---FYAVGTP--FLIFPALFPTVYIPLRDVLPTEPADR 128
DB 74 PLSTFIFS-LPVLSSKSLVLIHREFSLSLVCPILMMA-----SPHSNSK 119
QY 129 LQALIPGGL--GLVAIL---RANFPAFYVLALSGVMSLSM---FWGPAEITKI 178
DB 120 DAVEASGLTNDGANAATGMDNHGWFYISVRGFPLW-VALINLVAISSTWARIIDVMDI 178
QY 179 HEAKRFVALFGIGANISILASGRATVWASKLASVSEGVDPWGISLRLMAMTIVSGLVL 238
DB 179 QSGARLPFGVAGATL-----GQLFGSVPAATATW-MGPYLLPALMEPPA 225
QY 239 MASVYWIN-----KNVLTDRFYNPBEMOKGKGAKEPMNMK 275
DB 226 QSSKGTINDISQSSRELSPLRGTDNDHQERKQKTSKVSPP-----KVASFSPIS 278
QY 276 -----DSFLYDRSPYLLITLTV-----IAYGICINLIEVTKSQI-KLQ 315
DB 279 TTRDPFMAILQOMRLILASPYLLLSFLMLGAVISSEFFYFQKNIIATTIKSIGRRRL 338
QY 316 YPNMNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKPMGLGALT--PWVWLLGIYIF 373
DB 339 FAQINSF-----VAVFILIT--GQLT-----LGRILTVAGTVLALSASFV 377
QY 374 ALVIFRNOASGLVAMFGTTPMLAVVGAIQILSKSTKVALFDSTKEMAVIPLDQEQV 433
DB 378 AL-----GLVVAIAIMPTVTV--AVSETLAKVTYTVTRGRELLFTVSGDERY 426
QY 434 KKGKALIDVVAARFGSGGALLIOGGL-VICGSIGAMPTYLAVILLFIATIMLVSTKLNK 492
DB 427 KAKCIDIVIVORLGDAAA-----GLFEVLITLALGGQSTASIVALPVCILWITVAFPLGR 482

RESULT 20
08P142 PRELIMINARY; PRT; 453 AA.
ID 08P142
AC 08P142;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein XAC3056.
GN XAC3056.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RC MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farh C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camavayan F., Cardozo J., Chambergio F., Clapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorriy H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Guber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madela A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,

```

```

RA Spinoia L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RT Nature 417:459-463(2002).
DR EMBL: AE011948; AAA37901.1; -
KW Hypothetical protein, complete proteome.
SQ SEQUENCE 453 AA; 49247 MW; 34437007F15C2764 CRC64;

Query Match 7.2%; Score 188.5; DB 16; Length 453;
Best Local Similarity 20.6%; Pred. No. 7.9e-06;
Matches 107; Conservative 73; Mismatches 140; Indels 199; Gaps 25;

QY 31 FLMFPCITFNTVTLRDTKDLIVGAPSGAALPFIKFWLVPCAIIIMLYAKLSNLS 90
DB 28 FLVFPCLISGYVLRPVRAM--GA-SADVVAL-----FPTGMI----- 63
QY 91 KQALFYAVGTP-----FLIFPALFPTV--IYPLRDVLPTEPADRLQAI 132
DB 64 --AFPAAGVLPKQFTLOVLTCTFLIWLQPYGALVSRIPRX-----V 107
QY 133 LPPLGLGLVALIRWTFAPFYVLAE-----LWGSV---MLSLMFGPAEITKI 178
DB 108 FLPGYGFPIA-----TLLFVLPDSGVGRGMAFLWVTFNLPFAVAFMSFADVSN 163
QY 179 HEAKRFVALFG-----IG-ANISLSASG-----RAIIV-- 205
DB 164 AQAASYGYIGAAGTLGAFGLPVLTRVLRIGIAHMLVSAAGFLAVCVCLRILWAV 223
QY 206 ASKLKASVSEGVDPWGISLRLMAMTIVSGLVMAVSWYWINKNTVTDPRFNPBEMOKK 265
DB 224 AREBQGLSSEVP-----WGDVVGSLKLV----- 250
QY 266 KGAPKMMKDSFLYDRSPYLLITLTVIAYGICINLIEVTKSQI-KLQYPMNDYSE 324
DB 251 -----REPLRLWLAEMVL-FGVGVGLLNEQALVRLYTDAAAATA 292
QY 325 FMGNFSFWTGVSVLIMLFVGNVIRKPMGLGALT--PWVWLLGIYIF 384
DB 293 YSSSIDALINALVALQLVLRALISRFQ--IAPALLIPGVALIMGYALA-----ASP 344
QY 385 LVAMFGTTPMLAVVGAIQILSKSTKVALFDSTKEMAVIPLDQEQVKGKALIDVVA 444
DB 345 L-----PMMIAIV-----QVITRSSFPALAKAPARETLVTRVREWRKXGAALDVVY 392
QY 445 R-----FGKSGGALLIOGLVITCG-SIGA 467
DB 393 RCGDLTFWVWKLVSATGSS--AVFGVGLVATGTVGA 429

RESULT 21
084237 PRELIMINARY; PRT; 928 AA.
ID 084237
AC 084237;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE C. PSITTACI hypothetical protein.
GN CT234.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/Cx;
RC MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RT Science 282:1754-1759(1998).

```

DR EMBL: AE001296; AAC67826.1; --
 KW Complete proteome.
 SQ SEQUENCE 928 AA; 106121 MW; 8760C37E3D7CD75C CRC64;

Query Match 6.6%; Score 173.5; DB 16; Length 928;
 Best Local Similarity 19.9%; Pred. No. 0.00022;
 Matches 106; Conservative 88; Mismatches 195; Indels 143; Gaps 22;

7 KPGGLRS--FLMPHTHEKLVLMFELMFCITRYTVLRDTKTLVYGAGS-----58
 6 KGGSLRLRLRFLSLDRGEKRAL-LFLLGLI-----WSVACYGSLALGES 51
 59 -----GAELPPE-----IKFVLVPCALIFMLIYAKLSNLSKQALFYAVGTPLI-----104
 52 LPLEIGAEKLPALVIGASFL---CFISCLILYMLSRKRVSPKALFLSFICVLICNLX 108
 105 -FFALPFTVYPLRDVLPHTPEFADRLQALPPGLLGLVALILNMWTFAPFYLAELMG-SV 162
 109 LFMHL-----AIH-----KVGSGTPFEL-----YRLI--IMGLTI 136
 163 MLSLMFGFANETIKHEAKRYALFGIGANISLSAGAIYMAK-LRASVSEGVDPWG 221
 137 LCVANFPGFIDQFENIQAKRHCFI-----NATIFCGDFGARIVNOIQYLG 184
 222 ISRLIMMTIVSGVLMAASYWINKVLTDR-----FYNPEEMOKGKKAKPRMMK 275
 185 AELIILAFIVTTFPLVHYTSSLSKELSEBHNLELDGYRPTKQ-----TLKCLK 238
 276 DSFLYLDSPYLLTLTLVIAVGICINLIEVTWKSQKLQYNNM--DYSEFMGNPSFWT 333
 239 DKYTYLVSFYLMQLLVFT-----EFNYLKI FDAQFGNAETCELTENFTYSSWI 290
 334 GVVSVLIMLFVGNGVIRKFGMLTGALVTPVMVLIGIVFPALVIRNQGSLVANGTTP 393
 291 SLGNPFALFAYSRVITRG-----INNIILFAPICF-----FSLFCWS 330
 394 LMLAVVVAIQNILEKSTKVALFDSTKEMAYIPLDQEQVKAKAIDVVAARFGSGAL 453
 331 IKTSPFIATMGIAEGALVADDDNNQLIYGINRKIRNQRIAE-----SF 379
 454 IQQGLLVTCGSIAMTPYLAIVLIFTIAIWLVA---TKNKLFLAQSALK 501
 380 VEPAGMFICALCLFIPHOYVLCIIISAVCILLAILLRTHYSKAILRNLSTLE 431

RESULT 22

Q8NU46 PRELIMINARY; PRT; 538 AA.

AC Q8NU46; PRELIMINARY; PRT; 538 AA.
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Putative p-aminobenzoate transporter.
 GN CG10101.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 CC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
 CX NCBI_Taxid=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RL "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
 RT Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AP005274; BAB97494.1; --
 DR InterPro; IPR004697; Put transpt_Abgt.
 DR Pfam; PF03806; Abg_transport; 1.
 KW Complete proteome: Hypothetical protein.
 SQ SEQUENCE 538 AA; 57353 MW; 9D0E8465E7646B4 CRC64;

Query Match 6.2%; Score 162; DB 16; Length 538;
 Best Local Similarity 21.1%; Pred. No. 0.00083;
 Matches 127; Conservative 83; Mismatches 210; Indels 182; Gaps 29;

QY 2 TKTEEK-----PFGRLR-----SFLMPHT--HELKVLPMFLMFCITRYTVLRDTKDT 50
 DB 6 SKTSEKQPDPAFTSKLSKMSDKFLANGVETTLGNKLPFTPTLLILFLITALLASSSIAMNNV 65
 QY 51 LIVAGGS-----GAELPFIKMLVPCAI-----IFM 79
 DB 66 SVI-VPGSDEELFVKGFLPTBGLTWLTNGLNANYIGFPLTLVPIILAVGVAERSGMILA 124
 QY 80 LIYAKLSNLSKQALFYAVG---TPPLIFALPFTVYPLRDV-----HPTFPAURL 129
 DB 125 ALIRLTFGSARKIVLPVAVGVIGVASTADAAF-VVVPPLAAVFKAGRRHV-----177
 QY 130 QALIPGGLGLVAILRMTFAFYVLAELMGVMSLTMFGFANEI-----TKIHE 180
 DB 178 -----AGLGSFAV-----GAGYSTALV--PTSIDALFAGITNAVMTLPGIATTEVNP 225
 QY 181 AKRFYALFGIGANISL-LASGRAI-----VMASKLRASVSEGVDPWG-----ISRL 226
 DB 226 VSNY--FNIASIVLGLCGFLIDKVLPEPMWQKIATEYAESIEPTSAADDEISATL 283
 QY 227 -----LMMNTVSGVLMAASYWINKVLTDRFNFNPEEMOKGKKGAPK 271
 DB 284 TAOENRALTISMWTLTAPAILVLVVLIPGSPMREND-----GGFLPT 326
 QY 272 MNMDSFLYLDSPYLLTLTLVIAVGICINLIEVTWKSQKLQYNNNDYSEFMGNFSF 331
 DB 327 SPLSSVFT-----VLEPFMWGLAYGNVGTIK-----NMDDVVMNGEAL- 369
 QY 332 WTGVSVLIMLFVGNGVIRKFGMLTGALVTPVMVLIGIVFPALVIRNQSGL--VAMF 389
 DB 370 -KDMIGFLVLAIFLIGQFVALFNW-----TGIGWTAV---QGAAGLEAIGLT 412
 QY 380 GTPPLMLAVVGAION--ILKSTKVALFDSTKEMAYIPLDQEQVKAKAIDVVAARFG 447
 DB 413 GPALIAFIILASCLNLIISGSAMWTL-----MAAFVMPFALGYEPFIQAAFRVG 466
 QY 448 KSGGALI--QQGLLVTCGSIAMTP-----YLAIVLIFTIAIWLVSATKLNKLFQAOS 498
 DB 467 DSAQVITPLPNPVMVILIGLRRYEPDAGLGLMSRLIPFIPFLAATLILATIFYADL 526
 QY 499 AL 500
 DB 527 PL 528

RESULT 23

Q9PKG0 PRELIMINARY; PRT; 918 AA.

AC Q9PKG0; PRELIMINARY; PRT; 918 AA.
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical protein TC0505.
 GN TC0505.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 CX NCBI_Taxid=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MOPn / N159;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Base S.,
 RA Linher K., Weidman J., Knouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MOPn and Chlamydia
 pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL; AE002319; AAF39347.1; --
 DR TIGR; TC0505; --
 KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 918 AA; 105043 MW; 46D50ABF198D887 CRC64;
 Query Match 6.1%; Score 161; DB 16; Length 918;
 Best Local Similarity 19.0%; Pred. No. 0.0018;
 Matches 88; Conservative 80; Mismatches 198; Indels 96; Gaps 14;
 QY 57 GSGAEAIPIFKFWLVCACAIIFMLIYAK---LSNLIKQALFVAVGTPFLIFPALFPTVI 113
 DB 13 GEEKRALFLFLILGLVSVACGSLALGESLFEIEGKKLPFAVLGASFLFCF-ISCVL 71
 QY 114 YPLRDVLPTEFADRLQALIPGLGLVALIRNMTFAFYVLA-----EL 158
 DB 72 YSL-----SRKASPRALFLSFISCVLVCNLYLFWLAIHRGVSGPIFLYRILI 121
 QY 159 WG-SVWLSTLFWGANEITKHEKRYALFGIGANISLASGAIYMAK-LRASVSEG 216
 DB 122 WGLTILCYANFWGVDFQFNIDAKRHFCIF-----NAITFCGDFLGARLVNQ 169
 QY 217 VDPWGISLRLMAMTIVSGVLVMAVYMNKNVLTDDR-----FYNPEMOKGKGAAP 270
 DB 170 IOHLGAEILILSFAVIVFTFPLVHTYISSLSKEISEHDLELDYGYPPSAQAP----- 223
 QY 271 KMNKQDSFLYLDSPYILLLTLVYAYGICINLIEVTWKSQKLQYPMNDY--SEPMGN 328
 DB 224 KLCIKDKYTFYVSEFYFLQLLVFT-----EFNYLKI FDAQFGNAKPFELTENPTK 275
 QY 329 FSEFTGVSVILIMLVGAGNYTRKFGMLTGALVTPVMVLIGIVFALVIFRNQASGLVAM 388
 DB 276 YSSWISLGNMFALFAVSRVITRFG-----INNIIIFADICF-----FSL 315
 QY 389 FGTTPLMLAVVGAIQVILSKTKYALFDSFKEMAYIPLOEQKVKKAIDVVAARFGK 448
 DB 316 FVCKWISITSVIATWGMARGLAVLDNNMLQLIYIGIPKNINQVRIALE----- 367
 QY 449 SGGALIOQGLVTCGSGIGAMTPYLAIVLLFIATMLVSATRL 490
 DB 368 ---SFVEPAGMFIQALCLFPHQYIICIIISANCILLALL 406
 RESULT 24
 Q928P5 PRELIMINARY; PRT; 925 AA.
 AC Q928P5;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE CT234 hypothetical protein.
 GN CPN0293.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CWL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 DR EMBL: AE001614; AAD18442.1; -
 KW Complete proteome.
 SQ SEQUENCE 925 AA; 105615 MW; 98E6098E7C8FD37D CRC64;
 Query Match 6.0%; Score 158.5; DB 16; Length 925;
 Best Local Similarity 20.6%; Pred. No. 0.0027;
 Matches 95; Conservative 94; Mismatches 163; Indels 109; Gaps 26;
 QY 59 GAELPIPKFWL---VPCAIIFMLIYAKLSNLSKQALF-YAVGTPFLIFPALFPTVIY 114
 DB 49 GSAELP--KIVLGSSILICVLSLILVNLFKKHISATRALFIPVSLICNFIYILSSIF 106
 QY 115 PLRDVLPTEFADRLQALIPGLGLVALIRNMTFAFYVLAELMGVYML-SLFWGFRAN 173

DB 107 -----AIDPE-----RSLPF--FYRI-IMSLTILSYTFMGFVD 138
 QY 174 EITKIHAKRYALFGIGANISLASGRALVMAKLRASVSEGVDPKGISLRLMAMTIV 233
 DB 139 QFNLDGKHHCIFN-----AIIIFGDAL--OSGIIASLVHTIGIQGILILFTAAVLVT 191
 QY 234 SGVLVMAVYMNKNV--LTPDR--FYNPEMOKGKGAAPKKNMKDSFYLDSPYIL 288
 DB 192 FPLV-----FYVSKLSKLSDDHDLFD-----TGHPPLSKALKCFYDKTYRLL 238
 QY 289 -----LLTLVYAYGICINLIEVTWKSQKLQYPMNDY--SEPMGNFSTGVSVILIMT 342
 DB 239 CFYFLMQLAIA-----TEFNVLKIFELQFASKEFEIYAHIGKCSLWISGNMCFAL 291
 QY 343 FVGNVTRKFGMLTGALVTPVMVLIGIVFALVIFRNQASGLVAMGTTPLMLAVVGA 402
 DB 292 FAVSRVIRKLG-----VNNIIIFADICFSL-----LFLWTRKTLTSLIVLA 333
 QY 403 IONILSKTKYALFDSFKEMAY--IPLOEQKVKKAIDVVAARFGKSGALIQGLLV 460
 DB 334 M--VREGVTVYALDNNMLQLIYGVF---NKIRNQ--KIVVESFIEPIGMVMS---L 382
 QY 461 ICGSIGAMTPYLAIVLLFIATMLVSATRLKQLFQAQALK 501
 DB 383 IC--FLSSQGYVFCILISLIALTLVC---LVRSYVAKALIK 418

RESULT 25
 Q9J5I6 PRELIMINARY; PRT; 925 AA.
 AC Q9J5I6;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
 DE CT234 hypothetical protein.
 GN CP0293 OR CP0465.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uitterlidge T., Berry K., Baes S.,
 RA Linher K., Weidman J., Khouiri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.,
 RT "Genome sequences of Chlamydia trachomatis Moyn and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kunita S., Nakazawa T.,
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWL029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 DR EMBL: AP002208; AAR38302.1; -
 DR EMBL: AP002546; BAA98503.1; -
 DR TIGR: CP0465; -
 SQ SEQUENCE 925 AA; 105601 MW; 61E8941E7C8FD620 CRC64;
 Query Match 6.0%; Score 157.5; DB 16; Length 925;
 Best Local Similarity 20.4%; Pred. No. 0.0032;
 Matches 94; Conservative 95; Mismatches 163; Indels 109; Gaps 26;
 QY 59 GAELPIPKFWL---VPCAIIFMLIYAKLSNLSKQALF-YAVGTPFLIFPALFPTVIY 114
 DB 49 GSAELP--KIVLGSSILICVLSLILVNLFKKHISATRALFIPVSLICNFIYILSSIF 106

```

Oy 115 PLRDVLTHTPEBAPRLQALIPGLGLVALIRNMTFAFYVLAELMGVLS-LMFMGFAN 173
Db 107 -----AIDPF-----RSPLE--FYRIV-1MSLITLSTFSGFVD 138
Oy 174 EITKIHEAKRYALFGIANISLASGRAIYVASKLRASVEGVDPMGISRLMAMTIV 233
Db 139 QPFNLQDKRHFICFNM-----AIIPLGDAL--GSGIISLVHTIGIOGILLIFTLAVALT 191
Oy 234 SGLVIMASVMTNKNV--LTDP--FYNPEMOKKCAKPRMNMKDFVLNDRSPYL- 288
Db 192 PPIV-----FYVSKLSKLSDDHDLFID-----THPPLSKALKLCFYDKTFYLL 238
Oy 289 -----LTLVLVAYAGICINLIEVTWKSQKLOYPNNNDYS--EFMGNFSPMTGVSVLIML 342
Db 239 CFYFLMQLALTA-----TEFNLIKIFEIOPASKEEFELVAHIKCSLMTSLGMCPL 291
Oy 343 FVGNVIRKFGWLTGALVTPVWLLTGVIFPALVIFRNQASGLVAMFSTPLMLAVVGA 402
Db 292 FAYSRIVKRLG-----VNNIILFAPLCFLS-----LFLFWTFKTLTSLAVLA 333
Oy 403 IONLSKSTKXALPSTKEMAY--IPLDQOKVKKAIADVAAAFGSGGALLIOGSLV 460
Db 334 M--VREBVTYALDDNNLQLLIYGP--NKIRNQ--IRIVSSEFIEPIGLWVS--L 382
Oy 461 ICGSIGAMTPYLAIVLIFLIAIMLVSATKLNKFLAQSALK 501
Db 383 VC--FLSSQOYVFCIIISLIATILVC--LVRSYAKAILK 418

```

RESULT 26

O8YNE4 ID O8YNE4 PRELIMINARY; PRT; 1002 AA.

```

AC O8YNE4:
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein A114622.
GN A114622.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
CX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneke T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Nakazaki N., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Yaeuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003597; BAB76321.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1002 AA; 113795 MW; 4E060D76A2388BAE CRC64;

```

Query Match 5.7%; Score 151; DB 16; Length 1002;

Best Local Similarity 20.2%; Pred. No. 0.011; Mismatches 196; Indels 114; Gaps 23;

Matches 101; Conservative 88; Mismatches 196; Indels 114; Gaps 23;

```

Oy 19 IHTHELKVLPMFPCITFNVTYVLRDXTLLIVGAGSGAEAIPIFKFWLVVPCAIIF 78
Db 24 LRPEGERTWMMFAFYTVSGLRKADSTVALPLDEGAG--PLP-----WMTYASAVMG 77
Oy 79 M--LIYAKLSNLSKQALFYAVGTPFLIFPALPPTVYIPRDVLTHTPEBAPRLQALIP 135
Db 78 MALFVVSWMLOKIPFLR-----WVVAIAPCMIVPL-----ILLVLLRW 116
Oy 136 GL-----LGLVALIRNMTFAFYVLAELMGVLS-LMFMGFANITKIHEAKRYALFG 189
Db 117 GIDVAVYSVIVVFLRLMV--DSIYVNDLNTSIV-----ANQLENIRIKRTYPLIS 167
Oy 190 IGANISILASGRAIYVMA--SKLRASVEGVDPMGISRLM-----AMTIVSGLVIMASV 242

```

```

Db 168 SGLVADVIVSGFSLPMLLEBAKLN-----RYIMACGVIMIGSALICYLSY 213
Oy 243 WMTKAVLTDPFRFNPPEMOKKGAAPKXNMMDSPILDRSPYLLITLVLVAYGICIN 302
Db 214 QYRTSPFAPORLIPQOASRRRIQAP-----LKR--YTLQLFVGLLQITGL 261
Oy 303 LIEVTWKSQKLOYPNNND--YSEFMGNFSPMTGVSVLIMLIFYGVNIRKFGWLTGALV 360
Db 262 LVDFQYIQLKTI--NIGDEBLAGFLGIFGGIYGLCELTQWPFSSSLIERFGVFTFAL 318
Oy 361 TPVWLLTGVIFPALVIFRNQASGLVAMFSTPLMLAVVGAII-ONLSKSTKXALPDST 419
Db 319 LPVAV--GFVVPQMIIVLILPQIOS-----LAFPMGLVGLKFEDELRTYFVSS 367
Oy 420 KENAYITLDOEQKKGKALADVAAAFGSGGALLIOGSLVIGSIGAMTPYLAIVLIFT 479
Db 368 GPLLYQPI--PERLRSTQV-----LSGGT--AEAIATGTAG-----IIVITLTV 409
Oy 480 IAIMVSAATKLNKFLAQS 498
Db 410 CGLF-VPATMQKVFISGT 427

```

RESULT 27

O8TMZ1 ID O8TMZ1 PRELIMINARY; PRT; 533 AA.

```

AC O8TMZ1:
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Transmembrane efflux protein.
GN VARS OR MA2503.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
CX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atchoo D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwen P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grisham D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR EMBL; AE010943; AAM05889.1;
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub_transporter.
DR Pfam; PF00083; sugar_tr_1.
DR PROSITE; PSS0850; MFS; 1.
KW Complete proteome.
SQ SEQUENCE 533 AA; 56009 MW; A10C15C7FB89A34 CRC64;

```

Query Match 5.6%; Score 147; DB 17; Length 533;

Best Local Similarity 19.3%; Pred. No. 0.01; Mismatches 184; Indels 192; Gaps 21;

Matches 114; Conservative 83; Mismatches 184; Indels 192; Gaps 21;

```

Oy 23 ELKVLV-PMFMEFCITFNVTYVLR-----DRTDTLIVGAPSGAEAIPIFKFWLV 71
Db 2 ERTVLVALVLANFVLVIDTTIMVNSISALISPTNTVTVQQA-----ITLYAL 51
Oy 72 VPCAIIFMLIYAKLSNLSKQALF-----YAVGTPFLIFPALPPTV----- 112

```

Db 52 VMA5--FMITGGIGDIIIGKRAFLGLVYGVGS---LLTAISPTIGLVFVWMSILEGL 106
 Qy 113 ----IYPLRDVHPHPEFADRLQAIIPGLGLVALIRMTFAFYLAELMSWLSLMP 168
 Db 107 GATLWPAIQTIVTSNYEGEDRAI-AYGIIIGIV-----ASGIALGPIIGGMLTTAVT 158
 Qy 169 W--GPANETIKIHEAKRFYALFGIGANISILASGRAIYVASKIRASVSEGVDPWGISLRL 226
 Db 159 WRLNAGEVIVY-----IIVLALSRIIDLH-FLKAGEBEKLIVIGSILSA 202
 Qy 227 LMAMTVSGLVIMASY-WWINKNVLTDPFPYNEEMOKGKAKPRMKNDSFLYLDSP 285
 Db 203 LGMGLIVFGIIMAGTYGWW-----KARQTFSIAGIEISIP 236
 Qy 286 YLLLLTLVIAAGICINLEVTWKSQKLQYPMNDYSEFMGNFSGVSVLIMLPFG 345
 Db 237 FGLSPFPVFIAGVILLGFAAMERHL-----ISSGMPPLVR 273
 Qy 346 GNVRKFGMLTGLVTPVWVLLTGIVFALVIFRNQASGLVAM---FGTPIMLAVVGA 402
 Db 274 LDVLRDRVTSGLIQMWQTLFGGFLPSMALFLQIVGLNMQTGFTYLLPLSTPLIAS 333
 Qy 403 IONILSKSTKXALFDSTKE-----MAYIPLDOE----- 430
 Db 334 L-----TASRLSIFLASKRIIQAGVILLAGFLAIFATIDVEVRGLGLMTGFLIGIGG 388
 Qy 431 -----QYKKAALDVVAARSGSGGALIOGGLVIGSISAMTPYLAIV 474
 Db 389 LIASQVNNLVLSQVTPERISSETALMGTSQNLGMAIGTALMGLIAGLAVAGAIT----- 443
 Qy 475 ILLFIAT-----WLSATKLNKLEFLAQSALK 501
 Db 444 LIDSTALPEDLKPDLVSAVENOFLSNEIQ 476

RESULT 28

Q9KSH7 PRELIMINARY; PRT; 540 AA.
 AC Q9KSH7.
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Transporter, BCCF family.
 GN VCJ329.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RK MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolenko M.D., Vamathevan J., Baas S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Ueberback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae."
 RL Nature 406:477-483(2000).
 DR EMBL; AE004207; AAF94438.1; -.
 DR TIGR; VCJ329; -.
 DR InterPro; IPR000060; BCCF_transporter.
 DR Pfam; PF02028; BCCF_1.
 DR ProDom; PD01011; BCCF_transporter; 1.
 DR TIGRFAMs; TIGR00842; bccf; 1.
 DR PROSITE; PS01303; BCCF; 1.
 KW Complete proteome.
 SQ SEQUENCE 540 AA; 58464 MW; DAA6C5A49EE6B0E CRC64;

Query Match 5.5%; Score 145; DB 16; Length 540;

Best Local Similarity 20.5%; Pred. No. 0.015;
 Matches 106; Conservative 76; Mismatches 187; Indels 148; Gaps 24;

Qy 69 WLWPCAIIPMLI-----YAKL-----SNILSKQALFYAVGTPP-LIFALF 109
 Db 79 WLFWSGNLFIILFCVGLVISPFGKIRLGQAADKADHSFLSWLAMPFAAGMGIGLFWWSVA 138
 Qy 110 PTIVY-----PLRDVHPHPEFADRLQAIIPGL-LGLVALIRMTFAFYLAELMS 161
 Db 139 EPVAVYTWYETPL-----GVEANTPQAQLMGATMFGHLPW---AIYGV 183
 Qy 162 VMLSLMP-----MGFANEITIKIHEAKRFYALFGIGANISL-- 196
 Db 184 VALSLAFPCYKNGLEPLMSRVFYPILGGRANGACHVVDILAV--LATLFGATLSLGLGA 241
 Qy 197 -LASGRAIYVASKIRASVSEGVDPWGISLRLMAMTVSGLVIMASYWINKNVLTDP 254
 Db 242 QQAASGHHVFGFES-----GIGLQIAV-ICVVTGLATLSVVRGID----- 281
 Qy 255 FYNEEMOKGKAKPRMKNDSFLYLDSPFIILLTLVIAAGICINLEVTWKSQKL 314
 Db 282 -----GGVKVISNIMTVAFU-----LLALAIVIGYSTVLTLSIPTTMAVYLN 324
 Qy 315 QYPMNDYSE---FMGNFSP-----WTGVSVLIMLPFGGNVRKFGMLTGLVTP 362
 Db 325 LIPLSNPGRDEWVFGQWTFYVAMWISNSPFGMFLARVSKRTIRF--IYAVLIIP 382
 Qy 363 WMLLTGII-VEFALVIFR--NQASGLVAMFGTTPMLA-----VVVGAIONILSKSTK 412
 Db 383 TLVTLVWMSVFGGLAVDPVINEI-GVLQNGLTDSVLMFGQFDELVGKVLVIAVLYV 441
 Qy 413 YALPSTEMAVIPDO-----EQYKKAALDVVAARSGSGGALIOGGLVIT 461
 Db 442 LVFFITSSDGLVIDSITSGKLDSPVLYORFPAVAVAGAAVALLMIGTEAVQALO-- 499

RESULT 29

Q9JUS7 PRELIMINARY; PRT; 660 AA.
 AC Q9JUS7.
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Putative transmembrane transport protein.
 GN NMA1483.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
 RK MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
 RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrall B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis Z2491."
 RL Nature 404:502-506(2000).
 DR EMBL; AL162756; CAB84716.1; -.
 DR InterPro; IPR000060; BCCF_transporter.
 DR Pfam; PF02028; BCCF_1.
 DR ProDom; PD01011; BCCF_transporter; 1.
 DR TIGRFAMs; TIGR00842; bccf; 1.
 DR PROSITE; PS01303; BCCF; 1.
 KW Transmembrane; Complete proteome.

SQ SEQUENCE 660 AA; 74418 MW; 50046CF6DA3A37D1 CRC64;
 Query Match 5.5%; Score 144; DB 16; Length 660;
 Best Local Similarity 20.7%; Pred. No. 0.022;
 Matches 127; Conservative 99; Mismatches 171; Indels 216; Gaps 34;

QY 32 LMFPCITFNVTYLRDTKDTLIVGAPG-----SGAEAFPIKFK-WLVVPCALIFM----- 79
 DB 7 LFFVCVIV-----VLVLTVPDQVQWMLDRAKEVIFTEFSWFFVLTFSIFLGLLI 56
 QY 80 LIYAKLSNI-----LSKQALFYAVGTFP-LIFPALFPTVIYPLRDVHPTPE 125
 DB 57 LSVSSLGNIRLGRDEDEVEFGFLSWLMLPFAAGMGVGLMFVGAEPMLHMFSDITAGTPE 116
 QY 126 ADRLQALIPGLGLLVAIRNMTFAAFYVLAELMGVSLMPFWGAFNEITKIHEAKR-- 183
 DB 117 HRQOQAL-----HTFHWGVHAMS-----YGTIALALAYFGFR---YKLPALRSC 161
 QY 184 FYAL-----FGIGANI-SLASGRAIV-----WASKLRASVSE--GVDPWGISLRLL 227
 DB 162 FIPPLKEKISGRFGAIDIMALLATPFGIITTLGFGASQGLQGEIGMAENSFSVOYL 221
 QY 228 MAMTIVSGVLVMAVYMWINKVLTDPFRFYNPEMOKGKGA--PRQNMKDSFL---YL 281
 DB 222 IIAVMSLAIVSA-----ISGVKGKVKVLSLNLGLAFLLFFVL 261
 QY 282 DRSPIITLLTLVIAVYICI-----NLIEVTWKSQKLQYPMNDYSEFMGNFSF----- 331
 DB 262 AAGPVTYLLS-----AFGDNIGNYLGVLRLSFKT-----YAYERHKEKPFESWTVLYWAW 312
 QY 332 ---WTGVSVLIMLFVGNVIRKFGMLTGALVTPVWVLTGLVFPALVIFRQASGLVAM 388
 DB 313 WCSNAPFVGLFPIARISKRTIREF--VFGVLLIRG--LFGVLWF-----TV 354
 QY 389 FGTTPMLM-AVYVGAIONILSKSTKYALFDSKEMAYIPLDOEKV----- 433
 DB 355 FGNTAIWINDVAGVAGMLEKMTSSPETLIF--KFNFYLPPLPELTSIVSLVLSLFFVISA 411
 QY 434 -----KGKAA-----IDVVAARFGKSGGALIQGLLVICSGIGAM 468
 DB 412 DSGIYVNNITSRDKGLSAPRWQAVMWGVLSAVVLLMRSGG-----LGNLQSM 461
 QY 469 T-----PYLAVILFLIAIWL-VSA-----TKLN-----KLFLAQSALKEQ 503
 DB 462 TLIVSLPPLMLLMICFSLMKGLSADKKYFETRVNPTSVFTGKWKERLVQIMSQTOEQ 521
 QY 504 EV---AQEDSAPA 513
 DB 522 DILKFLKHTASPA 534

RESULT 30
 Q9JZ61 PRELIMINARY; PRT; 675 AA.
 AC Q9JZ61;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Transporter, BCCF family.
 GN NMB1277.
 OS Neisseria meningitidis (serogroup B).
 CC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 CC Neisseriaceae; Neisseria.
 OK NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / Serogroup B;
 RX MEDLINE=2015755; PubMed=10710307;
 RA Tetteijn H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Hickey R.J.,
 Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Dickson E.K.,
 Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 Mason T., Ciecko A., Parksey D.S., Blair E., Cifton H., Clark E.B.,

RA Cotton M.D., Utlarback T.R., Khouri H., Qin H., Vamathavan J.,
 RA Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
 RA Smith H.O., Frazer C.M., Moxon E.R., Rappuoli R., Venter J.C.,
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 RT MC58."
 RL Science 287:1809-1815 (2000).
 DR EMBL; AB002476; AAF1653.1; -.
 DR TIGR; NMB1277; -.
 DR InterPro; IPR000060; BCCF_transporter.
 DR Pfam; PF02028; BCCF; 1.
 DR ProDom; PD01011; BCCF_transporter; 1.
 DR TIGRFAMs; TIGR00842; bccf; 1.
 DR PROSITE; PS01303; BCCF; 1.
 KW Complete proteome.
 SQ SEQUENCE 675 AA; 76283 MW; 4BEC798C51D40196 CRC64;

Query Match 5.5%; Score 144; DB 16; Length 675;
 Best Local Similarity 20.7%; Pred. No. 0.022;
 Matches 127; Conservative 100; Mismatches 170; Indels 216; Gaps 34;

QY 32 LMFPCITFNVTYLRDTKDTLIVGAPG-----SGAEAFPIKFK-WLVVPCALIFM----- 79
 DB 22 LFFVCVIV-----VLVLTVPDQVQWMLDRAKEVIFTEFSWFFVLTFSIFLGLLI 71
 QY 80 LIYAKLSNI-----LSKQALFYAVGTFP-LIFPALFPTVIYPLRDVHPTPE 125
 DB 72 LSVSSLGNIRLGRDEDEVEFGFLSWLMLPFAAGMGVGLMFVGAEPMLHMFSDITAGTPE 131
 QY 126 ADRLQALIPGLGLLVAIRNMTFAAFYVLAELMGVSLMPFWGAFNEITKIHEAKR-- 183
 DB 132 HRQOQAL-----HTFHWGVHAMS-----YGTIALALAYFGFR---YKLPALRSC 176
 QY 184 FYAL-----FGIGANI-SLASGRAIV-----WASKLRASVSE--GVDPWGISLRLL 227
 DB 177 FIPPLKEKISGRFGAIDIMALLATPFGIITTLGFGASQGLQGEIGMAENSFSVOYL 236
 QY 228 MAMTIVSGVLVMAVYMWINKVLTDPFRFYNPEMOKGKGA--PRQNMKDSFL---YL 281
 DB 237 IIAVMSLAIVSA-----ISGVKGKVKVLSLNLGLAFLLFFVL 276
 QY 282 DRSPIITLLTLVIAVYICI-----NLIEVTWKSQKLQYPMNDYSEFMGNFSF----- 331
 DB 277 AAGPVTYLLS-----AFGDNIGNYLGVLRLSFKT-----YAYERHKEKPFESWTVLYWAW 327
 QY 332 ---WTGVSVLIMLFVGNVIRKFGMLTGALVTPVWVLTGLVFPALVIFRQASGLVAM 388
 DB 328 WCSNAPFVGLFPIARISKRTIREF--VFGVLLIRG--LFGVLWF-----TV 369
 QY 389 FGTTPMLM-AVYVGAIONILSKSTKYALFDSKEMAYIPLDOEKV----- 433
 DB 370 FGNTAIWINDVAGVAGMLEKMTSSPETLIF--KFNFYLPPLPELTSIVSLVLSLFFVISA 426
 QY 434 -----KGKAA-----IDVVAARFGKSGGALIQGLLVICSGIGAM 468
 DB 427 DSGIYVNNITSRDKGLSAPRWQAVMWGVLSAVVLLMRSGG-----LGNLQSM 476
 QY 469 T-----PYLAVILFLIAIWL-VSA-----TKLN-----KLFLAQSALKEQ 503
 DB 477 TLIVSLPPLMLLMICFSLMKGLSADKKYFETRVNPTSVFTGKWKERLVQIMSQTOEQ 536
 QY 504 EV---AQEDSAPA 513
 DB 537 DILKFLKHTASPA 549

RESULT 31
 Q8FUB5 PRELIMINARY; PRT; 563 AA.
 AC Q8FUB5;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Conserved hypothetical protein.

OS *Bacillus halodurans*.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 NC NCBL_Taxid=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=C-125;
 RM MEDLINE=99184645; PubMed=10086841;
 RA Takami H., Nakaone K., Hirama C., Takaki Y., Masui N., Fuji F.,
 Nakamura Y., Inoue A.,
 RT "An improved physical and genetic map of the genome of alkaliphilic
Bacillus sp. C-125."
 RL Extremophiles 3:21-28(1999).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL, AB013375; BAA75389.1; -
 DR InterPro, IPR004638; Efflux_EmrB.
 DR InterPro, IPR007114; MFS.
 DR InterPro, IPR005828; Sub_transporter.
 DR InterPro, IPR01411; TCR_TcR.
 DR Pfam, PF00083; augar_tf.1.
 DR PRINTS, PR01036; TCR_TcR.
 DR TIGRfam, TIGR00711; efflux_EmrB.1.
 DR PROSITE, PS50850; MFS.1.
 KM Transmembrane.
 SQ SEQUENCE 475 AA; 5136 MW; 1442844030A1BBE5 CRC64;

Query Match 5.3%; Score 140; DB 2; Length 475;
 Best Local Similarity 22.7%; Pred. No. 0.03; Indels 88; Gaps 22;
 Matches 110; Conservative 74; Mismatches 212;

QY 67 KFMVLVPCALIFMILYAKLSNLSKQAL-----FYAVGTPFLI-FFALPTVIYPLR 117
 DB 7 KKMVVV-CAVIFGSPFTMLNNNSMLNPAIPQMANVPEADAVATGVITTFVAMGMTPLT 65
 QY 118 DVLPHTPADRLQAILPPGLIGLVAIL--RWMTAAFTV---LAEIMGSV-----LSIME 168
 DB 66 GYL-GDKIKKEAYILIGIFVLSGLIGALSMNLPSTLVFPGLOGIGGGMPLSMTLIF 124
 QY 169 -----WGFANETK-----IHEAKRFYALFGIGANISLASGRAIVW 205
 DB 125 DAFPRNERGATGVWGVASMAPIPTGPGFIVETSNMXYLFVNIPFGLIGITIAVY 184
 QY 206 ASKLRAVSEGVDPWPGISL-----RLIMANTIVSGVLMAVWINKNVLTDPREFNPE 259
 DB 185 LPTKARSGQIKLDMNGFLFYAGVGSILLAGRMSDLTHLE--WINKVLLLLGALCLIV 242
 QY 260 EMQKGKGAQKPMNMKOSFLYDRSPYLLILLIVAYGI-CINLIEVTWKSQKLOYPN 318
 DB 243 FVQVEKRAEQPLD-----LSLFRIP--AYSLSIWVAGISSIGLFAIGIFVPLLLQ--Q 292
 QY 319 MNDSSEFMGNFSFMTGVSVLIMLFVGGNVIRKFGMLTGAALVTVPVW-----L 366
 DB 293 VYDGPMTGTGLVPLPSALFTGLTWSIGRMIDKRG-PSGINTAGMTIAAVGTFALGYIHL 351
 QY 367 LTGIVF-FALVIFNQASGLVAMFGTPTPLMAVVGAIONILSKSTKALPDKSTKEMAYI 425
 DB 352 ETGLMYIFAMWAIIVGVGMLTTPATTTGNMAIPEGILSRGSANNVLRQSSAFGIYFI 411
 QY 426 PLDDEQKVKGAIDVVAARFGSGGALLIQGLLVGSGISGAMTPYLAIVILLITIAVLV 485
 DB 412 SVPEFVR-RGQAL--VSTSFEEATLQAINEGFFI---VGFLT-----ALSDPAAYWLE 459
 QY 486 SATK 489
 DB 460 KKAK 463

RESULT 34
 020628 PRELIMINARY; PRT; 536 AA.
 AC 020628;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)

DE Cytochrome oxidase subunit 1 (EC 1.9.3.1) (Cytochrome c oxidase polypeptide I).
 GN COXI.
 OS Cythecocodium cohnii (Dinoflagellate).
 OG Mitochondrion.
 OC Eukaryota; Alveolata; Dinophyceae; Gonyaulacales; Cythecocodiaceae;
 OC Cythecocodium.
 NC NCBL_Taxid=2866;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WH-d;
 RM MEDLINE=97424386; PubMed=9280308;
 RA Norman J.E., Gray M.W.,
 RT "The cytochrome oxidase subunit 1 gene (cox1) from the dinoflagellate,
Cythecocodium cohnii."
 RL FEBS Lett. 413:333-338(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WH-d;
 RA Norman J.E., Gray M.W.,
 RT "A complex cox1 gene organization in the mitochondrial genome of the
 dinoflagellate, *Cythecocodium cohnii*: Homologous recombination
 generates two different open reading frames."
 RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO 1 IS THE
 CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 AND COPPER B (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME
 C + 2 H(2)O.
 CC -1- PATHWAY: RESPIRATORY CHAIN; TERMINAL STEP.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 DR EMBL, AF186994; AAB69658.1; -
 DR HSPSP, P18401; 1PFT.
 DR InterPro, IPR000883; COX1.
 DR Pfam, PF00115; CYCOXIDASE1.
 DR PRINTS, PR01165; CYCOXIDASE1.
 DR PROSITE, PS00077; COX1.1.
 KM Copper; Electron transport; Heme; Inner membrane; Membrane;
 KM Oxidoreductase; Respiratory chain; Transmembrane; Transport;
 KM Mitochondrion.
 SQ SEQUENCE 536 AA; 59328 MW; AA6BB4C821BD2C61 CRC64;

Query Match 5.2%; Score 137; DB 8; Length 536;
 Best Local Similarity 20.3%; Pred. No. 0.056;
 Matches 104; Conservative 73; Mismatches 174; Indels 162; Gaps 22;

QY 68 FWLIVPCALIFMILYAKLSNLSKQALFYAVGTPFLIFPALPTVIYPLDVLHPTPAD 127
 DB 81 FYLVMP-----GLFGFGNFYIVIFGSGSEVPVPP----- 109
 QY 128 RLQ--ALPPGLIGLVAIL-----RWMTF-----AAYVLA-BLMSYMLSLFWG 170
 DB 110 RVNNSFLILSLVLLILSLISEFGGCTWLYPPISTFPMNLSPSSIGITLGLISG 169
 QY 171 FANETIKHEAKRFYALFGIGANISLASGRAIVWASKLRASVSEGVDPWGI---SLRL 227
 DB 170 ISSVLTSLNFWITLISLRSIGITLKITS-----LFPMSLITSGML 211
 QY 228 MAMTIVSGVLMAVWINKNVLTDPREFNPEMQKGAQKPMNMKOSFLYDRSPY 286
 DB 212 LTLPLTGLALMI-LSDINVTLLFFDPIF-----GDPRIFYQHLFWFGHPEVY 259
 QY 287 TLIL-----TLIVAYGI-----CINLI-EVTWKSQKLOYPNNDSEF 325
 DB 260 ILIIPAGIISITISGLQKIIIFGNPSMIFAMSGISLGSVWGHMYTTGLGLEDTRSYF 319
 QY 326 MGNFSFMTGVSVLIMLFVGGNVIRKFGMLTGAALVTVPVWILT-----GIVFALVIFR 379

```

Db 320 SG-----VTLLISLPTGTKI---FNMWSTYVGNPSLLKTNSSLFGILLFL-LMFTI 367
Qy 380 NQASGLVAMGTTPLMLAV-----VGAIONILSKTKVALPDSKEMAYIPLDQEQVK 434
Db 368 GGSSTGII--IGNAAVDGLDHTYIIAHFHFVLSGAVIAIFSG-----IIFNIEKIG 419
Qy 435 GKAAIDVVAARFGKSGGALIQOGLVLCG-----SIG 466
Db 420 SKNLPSCSSNNRNVLVFTIGLITFGPMHFLGFVMPRRIPDPDSFISNFFLSIG 479
Qy 467 AMTPYLAIVLLFIITAIWLSATKLNKFLAOSA 499
Db 480 SGITLISFGFLFKNCWITSLAOC SRLFGHVA 512

RESULT 35
Q94Y38 PRELIMINARY; PRT; 577 AA.
ID 094Y38;
AC 094Y38;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2002 (TrEMBLrel. 19, Last sequence update)
DE Cytochrome oxidase subunit 1 (EC 1.9.3.1) (Cytochrome c oxidase polypeptide I).
GN COX1
OS Cythecodinium cohnii (Dinoflagellate).
OC Mitochondrion.
OC Eukaryota; Alveolata; Dinophyceae; Gonyaulacales; Cryptecodiniaceae;
OC Cryptecodinium.
OX NCBI_TaxID=2866;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WH-d;
RX MEDLINE=21531838; PubMed=11675595;
RA Norman J.E., Gray M.W.;
RT "A complex organization of the gene encoding cytochrome oxidase subunit 1 in the mitochondrial genome of the dinoflagellate, Cryptecodinium cohnii: Homologous recombination generates two different cox1 open reading frames.";
RL J. Mol. Evol. 53:351-363(2001).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO 1 IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICTOCHROME C + 2 H(2)O.
CC -1- PATHWAY: RESPIRATORY CHAIN; TERMINAL STEP.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY. EMBL: AF182643; AAI11630.1; -.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1.
DR PRINTS: PR01165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1.1.
KW Copper; Electron transport; Heme; Inner membrane; Membrane; Oxidoreductase; Respiratory chain; Transmembrane; Transport; Mitochondrion.
SQ SEQUENCE 577 AA; 64144 MW; FB8955EADCE78BFC CRC64;

Query Match 5.2%; Score 137; DB 8; Length 577;
Best Local Similarity 20.3%; Pred. No. 0.061;
Matches 104; Conservative 73; Mismatches 174; Indels 162; Gaps 22;

Qy 68 FMVLPVCAIIIMLIYAKSNLSKQALFYAVGTPLLPFLFPIVYIPLRDVLAHPTPAD 127
Db 122 FYLVMP-----GLFGGNGFYIYIFQSGPEVVP----- 150
Qy 128 RLQ--ALLPGLGLVAIL-----RNWTF-----AAFYVLA-ELMGSVMSLMPWG 170

```

```

Db 151 RVNPSILLISLSTVLLISLISIEFGCGTWTLYPPLSTFPMNLSPESIGNIIFGLISG 210
Qy 171 FANEITKIEAKFPYALFGANISLASGRAIWAASKLRASVSEGVDPGCI--SLRL 227
Db 211 ISSVLISLNFWITLISLSIGITLKTIS-----LFPMSLITSGML 252
Qy 228 MANTIVSGVLMASYMINKVNL-TDPRFYNPEEMQKKGAKPKMKMSFYLDSPY 286
Db 253 LTLPIITGLALMT-LSDINYNLTFPDI-----GGDIYQHLFWFGHPEVY 300
Qy 287 ILUL-----TLVIAVGI-----CINLI-EVTKSQDLQYPMNDYSEF 325
Db 301 ILIIPAFGIISIIISGLQKIIIFGNPSMTFAMCISILSGVWCHNYITGLFETRSYF 360
Qy 326 MGNFSEWTVGSVILMLFVGANVIRKGMULTGALVTPVMVLT-----GIVFALYIFR 379
Db 361 SG-----VTLLISLPTGTKI---FNMWSTYVGNPSLLKTNSSLFGILLFL-LMFTI 408
Qy 380 NQASGLVAMGTTPLMLAV-----VGAIONILSKTKVALPDSKEMAYIPLDQEQVK 434
Db 409 GGSSTGII--IGNAAVDGLDHTYIIAHFHFVLSGAVIAIFSG-----IIFNIEKIG 460
Qy 435 GKAAIDVVAARFGKSGGALIQOGLVLCG-----SIG 466
Db 461 SKNLPSCSSNNRNVLVFTIGLITFGPMHFLGFVMPRRIPDPDSFISNFFLSIG 520
Qy 467 AMTPYLAIVLLFIITAIWLSATKLNKFLAOSA 499
Db 521 SGITLISFGFLFKNCWITSLAOC SRLFGHVA 553

RESULT 36
Q03750 PRELIMINARY; PRT; 498 AA.
ID 03750;
AC 03750;
DT 01-JUN-1998 (TrEMBLrel. 05, Created)
DT 01-JUN-1998 (TrEMBLrel. 05, Last sequence update)
DE Cytochrome oxidase subunit 4.
GN NDH3.
OS Synechococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32049;
RN [1]
RP SEQUENCE FROM N.A.
RA Klughammer B., Sulemeyer D., Badger M.R., Price G.D.;
RT "Involvement of ndh3, ndh3 and ORP427 genes in high affinity CO2 uptake in the marine cyanobacterium, Synechococcus sp. strain PCC7002.";
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U97516; AAB62185.1; -.
DR InterPro: IPR003918; NADHUB Oxred4.
DR InterPro: IPR001750; Oxidored q1.
DR Pfam: PF00361; oxidored q1.1.
DR PRINTS: PR01437; NDOXDRDTHSE4.
KW NDH; Oxidoreductase; Plastocyanine.
SQ SEQUENCE 498 AA; 53900 MW; A0FBCD32A4AA8DE CRC64;

Query Match 5.2%; Score 136.5; DB 2; Length 498;
Best Local Similarity 20.6%; Pred. No. 0.056;
Matches 92; Conservative 60; Mismatches 139; Indels 155; Gaps 20;

Qy 82 YAKSNLSKQALFYAVGT-----PLIFPALPPTY-IYPLRDVLAHPTPADRLQALP 135
Db 55 YTFEHWMLSTIGNMYNVLGSLPLVINSLLTLVAIYISGRSNRPKLYSLLILINS 114
Qy 136 GLIGVAILBNWTPPAFYVLAELMGSVMSLMPWGANETTKTHKARFALRGIGANIS 195
Db 115 GITG--ALIAN-----NLLLFLEY-----ELIFYLLAI----- 145
Qy 196 LASGRAIWAASKLRASVSEGVDPWGISRLMLAMTIVSGVLMASYW--WINKVLTLD 252

```

```

Db 146 -----W-----GSEKKGYASTKFLIYTAISGLCVLAFLGIWVLSQSSFD 186
QY 253 PRFINPEEMOKKKAKPRKMMKDSFLYLDNSPYLLTLTLVYAGICINLIEV-TKMSQ 311
Db 187 FENLTLENIERTK-----VILLTLILIGIKIPIVPLHTW--- 223
QY 312 LKLOYPNNDYSEFMGNFSFWTGVVSVLIMLFVGNVIRKFGMLGALVTVPVMTLITGIV 371
Db 224 LPDAVVEANP-----AVTLL-----GGVPAKIG-----TYGLV 252
QY 372 FPAALVIFR--NQASGLVAMEGTTPLMLAVVGAIONILSKSTKXALPDSKEMAYIPLD 428
Db 253 RFGQLPDPVMSVSPALAVIGTVSMVGSILAIAIQRLDKRVAYS---SIGHMGYILVS 309
QY 429 Q-----BKKYKGAIDVVAARFGKSGALLIQ 456
Db 310 TAAGETSLGAVNAQMISHLALLFHLVGIIEKV-GRDLVDLNGIAMPVRLPLTS 368
QY 457 GLVLCGSGAMTP-----YLAIVILF 478
Db 369 SLTIAGMASGIPGLGVFAEFLVF 394

```

```

RESULT 37
Q8FH20 PRELIMINARY; PRT; 465 AA.
ID Q8FH20;
AC Q8FH20;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE Nitrite extrusion protein 1.
GN NARK OR C1684.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / ATCC 700928;
RX MEDLINE=22386234; Pubmed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.J., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Domeneberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AB016760; AAN80151.1; -.
KW Complete proteome.
SQ SEQUENCE 465 AA; 49877 MW; 73683DDB64CEFF5 CRC64;

```

```

Query Match 5.2%; Score 135.5; DB 16; Length 465;
Best Local Similarity 22.5%; Pred. No. 0.062;
Matches 109; Conservative 68; Mismatches 174; Indels 133; Gaps 28;
QY 69 WLVPICAIIFMLIYAKLSNI-LSKQALFYAVGTPFLIEFALFPYTYPLRVLHPTREAD 127
Db 41 WISPCLLIACVWMLBSAVAVNLEPKVFNFTDQLFMTLALPSVSGALLV--PYSF-- 96
QY 128 RLQALIPPLGLGLVAIL--RNMTFAAFYLAELMGSVMLSLMFGFANEITKIHAKRFY 185
Db 97 -----WPIPGGRMT--AFST-----GILLIPCVWJGFA-----VDTSIRPY 132
QY 186 ALFGIGANISLASGRAIWMASKLRASVS-----EGVDPWGISRLMLAM 230
Db 133 SVFTI--ISLLCGFAGANFASSM-ANISFFPKQKOGALGLNGIGNMGVSVWQVLAP 188
QY 231 TIVGVIVMASYWMINKVVLDPFRFYNPEEMOKKKAKPRKMMKDSFLYLDNSPYLL-- 288
Db 189 LVVS-LSPFAVF-----GSQGVKQPDGTE--LYLANASWIVP 223
QY 289 LTLTLVYAGICINLIEVTKSOLQLQYNNMND-----YSEFMGNF-SFWTGV--- 335

```

```

Db 224 LLAIFTIAMEGNMDL-ATSKASIXEQLPVLRGHLWIMSLLYLATFGSFIGSAGFAML 282
QY 336 -----VSUILMFVG---GNVIRKFGW-----LTGALTVPV---MVLITGVFPALV 376
Db 283 SKTQFPDVOILQVAFSPFIFGALARSAGALSDRLGSTRVTLVNFILMAIFSGLLFTLP 342
QY 377 IFRQASGLVAMEGTTPLMLAVV-----GAIONISKSTKXALPDSKEMAYIPDOE 430
Db 343 T-DQGGSPFAAFFA--VFLLALPLITAGSGSTFGOMSVIFRKLTMRYAEG---GSD 394
QY 431 QKVYKGAIDVVA-----RFGKSGALLIOGL--LVIGGS-IGAMTPYLAIVILPLIA 481
Db 395 ERARERATDPAALGFTSAIGAIGFPIPAFGSSIALTGSFVGAMKVFILFIYACVI 454
QY 482 INLV 485
Db 455 TWAV 458

```

```

RESULT 38
Q916P7 PRELIMINARY; PRT; 441 AA.
ID Q916P7;
AC Q916P7;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Probable MFS transporter.
GN PA0241.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; Pubmed=10984043;
RA Stever C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou Y.,
RA Garber R.L., Goltzy L., Tolentino F., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A.R., Larbig K., Lim R.M.,
RA Smith K.A., Slater M.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J.A., Siler M.D., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AB004462; AAG03630.1; -.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub. transporter.
DR Pfam; PF00083; sugar tr; 1.
DR PROSITE; PS50850; MFS; 1.
KW Complete proteome.
SQ SEQUENCE 441 AA; 48704 MW; 72EDFF82986FB601 CRC64;

```

```

Query Match 5.1%; Score 135; DB 16; Length 441;
Best Local Similarity 21.1%; Pred. No. 0.063;
Matches 110; Conservative 58; Mismatches 163; Indels 190; Gaps 26;
QY 32 LMFPCITFNTVLRDTCITIVG-----APSGAEKIPF--IKFVLVPCAIIFMLIYA 83
Db 39 LVFPAATLNI-----DRAALGVMPILAEKMSWTAMDYANINFWQVAYAGFIQGR 92
QY 84 KLSNLSKQALFYAVGTPFLIFALFPYTYPLRVLHPTREADRLQALIPGLGLVAI 143
Db 93 FIDKGVKRAFPFLAV-----LMSLATGAGLATS 122
QY 144 LRMTFAAFV-IAEL--MGSVMLSLMFGFANEITKIHAKRFY--LFGIGANISILA 198
Db 123 AAGFMVORFLIGLEAANYACVKTIRLWPPAGE-----RAVAGINAGANNVAGAV 174
QY 199 SGRAI-----VMASKLRASVSEGVDPWGISRLMLAMTIVSGVLNASYWMINKVITDP 253
Db 175 TPALLPLILGVWG-----WQAAFLCMSAL--GLVWLV-FWWRN----- 209

```

```

QY 254 RFYNPEEMOKKKG-----AKPKNMKDSFLYDRSPYLLTLVIA----- 296
D 210 -YNPEEHRVYKQSLLEYIQOEBPPATRVPSQILRRGTMARFALASITAPVFWFLY 268
QY 297 -----YGCINLEVTWKSQKLOYPNNNDYSEFMGNFMTGVSVLIML----- 342
D 269 WLPPFLNOQYIGISVTCM-----GIFLLIMLTDGFC 301
QY 343 FVGNNVRKFGMLTGVLPVWVLLTGIVFAL-----YIFRNQASGL-VANFGTTPMLA 397
D 302 SVGGGILS--SMLGRGMPARLRLSMLLPACTIVGVAFANASGLMIAV-----LAIA 354
QY 398 VVGAIO-----NLSKSTKVALFDSTKEMAYIPLDOEKVKAKAIDVVAARFG-----KS 449
D 355 LAVGHHQMTANINS-----LYMDTTPKHLMTSTVG-----FGCMCAL 393
QY 450 GGALIOGGLVTCGSIGAMTPYLAVILFIITIMLVSATKL 490
D 394 GGMFTQ---IVGVLTATNNNAVLFMTIPMAYFIALTWL 431

RESULT 39
QY 08G4F6 PRELIMINARY; PRT; 404 AA.
AC 08G4F6;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Hypothetical protein.
GN Bli430.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteriales; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22949477; PubMed=12381787;
RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA Persi G., Zakhlen M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
KM EMBL: AF014773; AAN5227.1;-.
KN Hypothetical protein; Complete proteome.
SQ SEQUENCE 404 AA; 43122 MW; A6D16CC67EC251B7 CRC64;

Query Match 5.1%; Score 134.5; DB 16; Length 404;
Best Local Similarity 20.9%; Pred. No. 0.062;
Matches 92; Conservative 74; Mismatches 168; Indels 107; Gaps 24;

QY 95 FYAVGTPPLFFALPFTVYPLRDVLPTEFADRLQAILP-----PG 136
D 9 FISGLDPSILSGAMPTW---SODLNVAVAMAGISAVISMFTIVSALLSDRMTLKFGAG 65
QY 137 LLEGVALI-LRWMTFAAFVYLAELMGVSLMFP-WGFA-----NEITIHAKR--- 183
D 66 KVIVASVALTAALAGSVTSNYM--VLLAIAIPYGLAGVDALNNYVAIHESRIMS 123
QY 184 -FYALFGANISILASGRATVWASKLASVSEGV--PMGIS-----LRLMAMTIVSGLV 237
D 124 WILHGMGVGASVGYIMGYAL-----SQGGGPMGFRYRIALLQVMLTVLVSLP 173
QY 238 LMASYMNKNVLTDPREFYNEEMOKKGAAPKNNKDSFLYDRSPY-----ILLT--- 290
D 174 LMKRGAIAVSESTD-----TASDGAERGTAEVSVARERKPLGVAVLAIRGA 225
QY 291 -TLIVAVGICINLIEVT---WKSQKLOYPNNNDYSEFMGNFMTGVSVLIMLFPVG 346
D 226 KEILVMPF--CYCAVESTAGIMASSYMWHSIGDIKTA---ASWAS-----LFYGI 272
QY 347 NVIRKF--GMLTGALTVPVWVLLTGIVFALVIFRNQASGLVAMFGTTPMLAVVGAION 405

```

```

D 273 TVGRALSGFLMRPKDPFMIILGQVLVA-----GILTMVPLPHLHGVLVGLVI 323
QY 406 ILSKSTKY-ALPDSKEMAYIPLDOEKVKAKAIDVVAARFGSGGALIQ---GLLVI 461
D 324 GFGCAPYPCVHSTP--AYGEBDSQAIYG---VOMACAVV-----GSLMPPLFGIIAQ 374
QY 462 CGSIGAMTPYLAVILFIITAI 482
D 375 YATISLPWVLLVLLVLMVAM 395

RESULT 40
QY 09V250 PRELIMINARY; PRT; 976 AA.
AC 09V250;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Oligonucleotidyl transferase.
GN PA82202.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RA Hellig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ248283; CAB49148.1;-.
DR InterPro: IPR003674; OLIgo_trans_STT3.
DR Pfam: PF02516; STT3; 1.
KM Complete proteome.
SQ SEQUENCE 976 AA; 109052 MW; D27A162C72E51A68 CRC64;

Query Match 5.1%; Score 134; DB 17; Length 976;
Best Local Similarity 20.9%; Pred. No. 0.18;
Matches 107; Conservative 79; Mismatches 157; Indels 168; Gaps 29;

QY 59 GAEAIPIKFWLVVPCAIIFLVIYA-----KLSNLSQAL--F 95
D 19 GKSILYPLK-RILPLAVIGGIIAYLYRHLTACKYRPDPPTFHFETIKLVKEGLPKY 77
QY 96 Y-----AVGTPPLFFALPFTVYPLRDVLPTEFADRLQAILP--PP--GLIGIVA 142
D 78 YPMAEAPFGSLIGBPLGY--ILPAIFYKVVSFGVNEF---QAFIMMPFVFGIYIA 131
QY 143 -----ILRWTFFAAFVYLAELMGVSLMFGFANEITKIHAKRFA--LFGIGA 192
D 132 VYLLGRVYLNEM-----AGLMAAVILSVSTANFRTSGNARGDGPFTLPLFSIVA 183
QY 193 NISLSAGS--RAIVMAS--KLRAVSSEGV--DPMGISRLMAMTIVSGVLMASVW 244
D 184 MLYTLKENDITKKSILMAVAVFLASISLGAANGSPFGLM-----LIPASQOT 232
QY 245 INKNVLTDPREFYNEEMOKKGAAPKNNKDSFLYDRSPYLLTLVIAVIGICINLI 304
D 233 IALPI-----FGKIKELKKFKERYP-----AYLAIAGVGLTI--- 267
QY 305 EVTWKSQKLOYPNNNDYSEFMGNFSE--WTGVSVLIML.FVGNVIRKGMLTGALVTP 362
D 268 -----PGIAIKGFI--KFAFEVPLGLVLTIMLYGKFLNYSDKGRFAVVA 314
QY 363 VMVLTGIVFALVIFRNQASGLVAMFGTTPMLAVVGAIONLSSTKVALDSTKEM 422
D 315 VIVILG---FA-----GAYAYVG--PKLFRIMGAYQ---STQ--VYQVDEL 352
QY 423 AYIPLDOEKVKAKAIDVVAARFGSGG---ALIOGGLVTCG-----SIGAMT 469
D 353 A-----KTTLSIDIKLYYGVGNNGLVFFLSIPGILLILGLYNALLKKSSSN 400

```


Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamiales; Lamiales; Orobanchaceae; Rhinanthaceae; Triphysaria.

REFERENCE
1 (bases 1 to 966)

AUTHORS
Matvienko, M., Torres, M.J. and Yoder, J.I.

TITLE
Transcriptional responses in the hemiparasitic plant *Triphysaria versicolor* to host plant signals

JOURNAL
Plant Physiol. 127 (1), 272-282 (2001)

COMMENT
21437952
11553755

CONTACT: Yoder, J.I.
John I. Yoder Research Lab, Dept. of Vegetable Crops
University of California at Davis
137 Hammond Hall, One Shields Drive, Davis, CA 95616, USA
Tel: 530 752 1741
Fax: 530 752 9659
Email: jiyoder@ucdavis.edu
length = 966 bp.

FEATURES
SOURCE
Location/Qualifiers

1..966
/organism="Triphysaria versicolor"
/mol_type="mRNA"
/db_xref="taxon:64093"
/issue_type="root-tips ~5mm length"
/dev_stage="3-4 weeks growth"
/lab_host="E. coli"
/clone_lib="Triphysaria versicolor root-tip, early
DMBQ-induced transcript cDNA library"
/note="Vector: pCR2.1 TA Cloning System, Invitrogen,
Carlsbad, CA; PCR-based suppression subtractive
hybridization cDNA library"

BASE COUNT
222 a 186 c 240 g 318 t

ORIGIN

Alignment Scores:

Pred. No.: 1..49e-81 Length: 966
Score: 842.50 Matches: 163
Percent Similarity: 68.09% Conservative: 61
Best Local Similarity: 49.54% Mismatches: 96
Query Match: 32.03% Gaps: 9
DB: 10 Indels: 2

US-09-869-433-2 (1-515) x BE574942 (1-966)

QY 84 LysleuSeranlleuSerlyeGlnAlaLeuPheTyAlaValGlyThrProPheLeu 103
DB 6 AAGTGGCTAATGTTTGTGCAAGATGACCTTTCTACACTGTTATGTTCCATTCAT 65
QY 104 lIlePhePheAlaLeuPheProThValIleTyProleuMaGaPaValLeuHieProTh 123
DB 66 GCGTTTGGGGCATTGGGTTGGTTTATGCTTCAGTAATTTTTCACCCCACT 125
QY 124 GluPheAlaAsPaRyGleGlnAlaIleuPProFroGlyLeuGlyLeuValAlaIle 143
DB 126 GCGCTTCTATAGCTTTTGAAGGTGGGTCGAGGTTCTTGACCACTTGCTATT 185
QY 144 LeuArgAnThrPheAlaAlaPheTyValLeuAlaGluLeuTyPglySerValMet 163
DB 186 TTGAGGATTGGACCTTTGTTGTTCTATGTTATGCTGAGCTTGGGCGAGTGGTT 245
QY 164 LeuSerLeuMetPheTyPglyPheAlaAsnGluIleThyAlaIleHieGlnAlaValArg 183
DB 246 GTTCTGTTGTTGTTGGGTTTGGCAATCAATCAACCAATCAACCAAGCAAGAA 305
QY 184 PheTyAlaLeuPheGlyIleGlyAlaAsnIleSerLeuAlaSerGlyAlaAlaIle 203
DB 306 TTCTATCCTCTTTTCGGCTCGAGCAATGATGACACTTCTCTCGGTCGAGCTG 365
QY 204 ValTrpAlaSerLySLeuArgAlaSerValSerGluGlyValAspProTyPglyIleSer 223
DB 366 AATACTTTTCCCAATGAGCAAAATTTGGCCCGGTTGATGTTGGGCGCATCTCG 425

QY 224 LeuArgLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAlaSerTyTrp 243
DB 426 TTGAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 485
QY 244 TrpIleAnLySAnValLeuThAspProArgPheTyTrsAnProGluGluMetGlnLyS 263
DB 486 TGGGTAAACAAT-----TATGTTCTCTTCGACCCGAGT 521
QY 264 GlyLySylGlyAlaLySProLySMeTAsMeLySAsPsePheLeuTyTrleuAsPaRg 283
DB 522 ATGAAGAAG 581
QY 284 SerProTyTrleuLeuLeuThLeuThLeuValIleAlaTyRgLyIleQySIlleAnLeu 303
DB 582 AAGAAATATATTAGATTTTGGCCACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 641
QY 304 lIleGluValThrTrpLySserGlnLeuLySleuGlnTyRProAnMeTAsnAspTySer 323
DB 642 GTGAGGTTACATGAAATCAAGACTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCA 701
QY 324 GluPheMetGlyAnPheSerPheTyRgTyAlaValSerValIleuIleMetLeuPhe 343
DB 702 GCTTCATGAGGTGCTTTCAACTACTACCGGA---ATAGCACTTCAATATATGCTT 758
QY 344 ValGlyLySAnValIleArgLySPhelGlyTyRleuThRgTyAlaLeuValThProVal 363
DB 759 TTGAGTCAATGATATTCACCAATACAGTTGGGAATGGCAGCAAGATTAACCGACC 818
QY 364 MetValLeuLeuThRgTyIleValPhePheAlaLeuValIlePheArgAnGlnAlaSer 383
DB 819 GTCTGCTAATCACTGGGATGCTTCTTCTCGTGTATCTTTTCGGAGACCTTTGGCC 878
QY 384 GlyLeuValAlaMetPheGlyThThProleuMetLeuAlaValAlaIle 403
DB 879 CCGTCTCGCCCGCTTCGAAATGACCTCTATTAACAGCTGTCTATGTTGGGCGAGT 938
QY 404 GluAnlleuSerLySserThrLyS 412
DB 939 CAATACTTTTATGTAAGAGCGCAAG 965
RESULT 3
CB644624
LOCUS
DEFINITION
OSUNEB05020.f OSUNEB Oryza sativa (japonica cultivar-group) cDNA
clone OSUNEB05020.5', mRNA sequence.
ACCESSION
CB644624
VERSION
CB644624.1 GI:29639615
SOURCE
ORYZA
ORYZA sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
1 (bases 1 to 822)
Xanthoxanthinase; Oryzae; Oryza.
AUTHORS
Kudrna, D., Dean, R., Soderlund, C., Wang, R. and Wang, G.
TITLE
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
JOURNAL
Unpublished
CONTACT: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR PRIMERS
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 05 row: O column: 20
Seq primer: gta aac cga cgg cca gtc.
Location/Qualifiers

FEATURES

| | source | 1.. 822 |
|--|--|---------|
| | Organism="Oryza sativa (japonica cultivar-group)" | |
| | /mol_type="mRNA" | |
| | /culivar="Nipponbare" | |
| | /db_xref="taxon:39947" | |
| | /clone="OSUNEB05020" | |
| | /tissue_type="Leaf" | |
| | /dev_stage="3 week" | |
| | /lab_host="DH10B" | |
| | /clone_id="OSUNEB" | |
| | /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2: XhoI; 24 hrs after inoculation with Rice Blast (Cne 86061)" | |
| BASE COUNT | 190 a 166 c 231 g 235 t | |
| ORIGIN | | |
| Alignment Scores: | | |
| Prod. No.: | 8.87e-71 | 822 |
| Score: | 743.50 | 142 |
| Percent Similarity: | 71.22% | 56 |
| Best Local Similarity: | 51.08% | 75 |
| Query Match: | 28.27% | 5 |
| DB: | 14 Gaps: | 2 |
| US-09-869-433-2 (1-515) x CB644624 (1-822) | | |
| Oy | 147 TTPThrPheAlaAlaPheTyValLeuAlaGluLeuTTPGlySerValMetLeuSerLeu | 166 |
| Db | 2 TGGACCTCTGCCTCTTCTATGTCATGCGCCAGCTCTGGGGCAGCGTGCATCTCCGG | 61 |
| Oy | 167 MetPheTPGlyPheAlaAsnGluLeuThrLysIleHisGluAlaLysPhePheTyAla | 186 |
| Db | 62 CTCCTTGGGGATTGCCATCAATCAATCTACGTTGAAGAGCCAAAGAGTTTACCA | 121 |
| Oy | 187 LeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGlyAlaAlaIleValTTPAla | 206 |
| Db | 122 CTGTTGGGGCTTGGGCCCATGTGCACTCATCTTCTGGGTCGACAGTGAATACCTTC | 181 |
| Oy | 207 SerLysLeuAlaGalaSerValSerGluGlyValAspProTTPGlyIleSerLeuArgLeu | 226 |
| Db | 182 TCMAACATGAGGCGAATTTGGGTCCAGGGGTGAGCGGATGGCGATTTCATTGAGGGGC | 241 |
| Oy | 227 LeuMetAlaMetThrIleValSerGlyLeuValLeuMetAlaSerTyTTPTriPheAsn | 246 |
| Db | 242 ATGATGACATAGTGCTTTGCTGCGCCCTTGATGATCCCGCATCTACTGGGAGTAAC | 301 |
| Oy | 247 LysAsnValLeuThrAspProArgPheTyAsnProGluMetGlnLysGlyLysLys | 266 |
| Db | 302 AAGTTTGTAAATTTGATAAA-----TCAGCTGGCGCCAGTGTGGAGGAAGAAG | 349 |
| Oy | 267 GlyAlaLysProLysMetCAsnMetLysAspSerPheLeuTyLeuAspArgSerProTy | 286 |
| Db | 350 AAGGATTAAGCCAAAGCTGAATGGGAGAGAGAGTTGAAGTGCTGTCTCATCTCGTAT | 409 |
| Oy | 287 IleLeuLeuLeuThrLeuLeuValIleAlaTyGlyIleCysIleAsnLeuIleGluVal | 306 |
| Db | 410 GTGAGGAGATCTTGGCCACCTGCTGCTGCTTATGTTATGATTAACATTAACCTTGAGG | 469 |
| Oy | 307 ThrTTPLysSerGlnLeuLysLeuGlnTyProAsnMetAsnAspTySerGluPheMet | 326 |
| Db | 470 ACATGGAAGTCAAAACTGAAGGCGCAAAATACCACCCCTTAAGTAATATCTTCATTCATG | 529 |
| Oy | 327 GlyAsnPheSerPheTTPThrGlyValAlaValSerValIleLeuMetLeuPheValGly | 346 |
| Db | 530 GGGGATTTCTGCACCTGCTACTGCTAGCGCATGCGCATTTTACATGATGATG--TTAGAGGA | 586 |
| Oy | 347 AsnValIleArgLysPheGlyTTPLeuThrGlyAlaLeuValThrProValMetValLeu | 366 |
| Db | 587 GTAATTCCTTAAGAAATTTGGGGGAGGCGGACGCTAATCAATCAACCCCTGCAGTGTCCT | 646 |
| Oy | 367 LeuThrGlyIleValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuVal | 386 |
| Db | 647 CTCACGTGAAGTTGATTTCTTCACCTGATTTATTCGGGTGAGCACTTCACTTCCTTATG | 706 |

```

OY 387 AlameetPhegiLYTrThProlemetLeuAlaValAlaGlyAlaileglnsnnle 406
DB 707 GCCACACTGGAATGACACCTTGTCTCTCTCGCGCTCATGTTGGGCGCATGCAAACTT 766

OY 407 LeuSerLYSerThrLYeTYrAlaLeuDehApsSerThrLYeGLMeLaLYr 424
DB 767 TTCAGTAGAGTGGCAAGTACAGTTGTTTACCTTGTAAAGAAATGSCATAC 820

RESULT 4
DB CB655122 724 bp mRNA linear EST 09-APR-2001
LOCUS OSJNEC08D15.f OSJNEC Oryza sativa (japonica cultivar-group) cDNA
DEFINITION clone OSJNEC08D15 5', mRNA sequence.
ACCESSION CB655122
VERSION CB655122.1 GI:29658847
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehretidoideae; Oryzae; Oryza.
1 (bases 1 to 724)
JantaseurLYar,C', Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wang,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished
JOURNAL Contact: Rod Wing
COMMENT Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 08 row: D column: 15
Seq primer: gta aaa cga cgg cca gtc.
Location/Qualifiers
1..724
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEC08D15"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEC"
/notes="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2
XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"

BASE COUNT 128 a 231 c 194 g 171 t

ORIGIN

Alignment Scores:
Pred. No.: 1,33e-68 Length: 724
Score: 723.00 Matches: 131
Percent Similarity: 75.43% Conservative: 44
Best Local Similarity: 56.47% Mismatches: 57
Query Match: 27.49% Indels: 0
Gaps: 0

US-09-869-433-2 (1-515) x CB655122 (1-724)

OY 19 lIehsThrnIsGtLeuLYsLYsValLeuBrometPheLeuMetPhePheCyslleThr 38
DB 21 GTGAGACGAAGACGCTGAGAAAGATGTCGCGCTGGGGGCTCATGTTCTTCGCACTTC 80

OY 39 PheanTYrThrValLeuLYAspThrLYAspThrLeuileValAlaIAspProLYser 58
DB 707 GCCACACTGGAATGACACCTTGTCTCTCTCGCGCTCATGTTGGGCGCATGCAAACTT 766

```

Db 81 TTCACATACACATCTGCGGACACCAAGACGTGCTGCTACCGCCCAAGGCGAC 140
 QY 59 GYALAGLALALEPHEHELLEYSPEHETPLLEVALVALPROCYALALELLEPHE 78
 Db 141 ACGCGGAGATATCCCGTTCTCCACAGAGCGTGGGTAACTCCCAATGCGCATCGGCTTC 200
 QY 79 MetLeuLeuTyrAlaLysLeuSerLeuLeuSerLeuSerLeuValLeuPheTyrAlaVal 98
 Db 201 ATGCTGCTTACACCAAGCTCCGCCAGCTGCTTCCAGAGGCCCTCTTACACCGCTC 260
 QY 99 GlyThrProPheLeuLeuPhePheAlaLeuPheProThrValLeuTyrProLeuArgAsp 118
 Db 261 ATCTTCCCTTCATCCCTTCTTCCGCGCTTCGAGCTTCTGCTTACCTTACCCCATGCGCGAC 320
 QY 119 VALLEUHLISPROTHGLUHEALAEAPARGLEUGLNALILEUENPROGLYLEUEN 138
 Db 321 GCCATCCACCCACCGCGCTCGCGACCGCTCTTCGCGCGCTGGGCGCCAGCTTCCG 380
 QY 139 GYLeuValAlaLeuLeuArgAsnTyrThrPheAlaAlaPheTyrValLeuAlaGluLeu 158
 Db 381 GCGCGCGTCCATCTCCGCGCTCGAGCTTCTGCTTCTTATGTCTATGGCCAGCTC 440
 QY 159 TTPGlySerValMetLeuSerLeuMetPheTTPGlyPheAlaAsnGluLeuThrTyrSile 178
 Db 441 TGGGCGACAGCGTCTCATCTCCGTGCTCTTGGGAGATTGCGCAATCATGATTACTACGCTT 500
 QY 179 HIsGluAlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnLysSerLeuLeuAla 198
 Db 501 GAAGAGGCGCAAGAGTTTACCACCTGTGGCGCTGGGCGCAATGTGCACTCATCTTC 560
 QY 199 SerGlyArgAlaIleValTrrpAlaSerLysLeuArgAlaSerValSerGluGlyValAlaP 218
 Db 561 TCGGCTCGACAGTAAATCTTCTCAACATGAGGACAGATTGGCTCCAGGGGTGAC 620
 QY 219 ProTTPGlyLysSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeu 238
 Db 621 CGATGGCGCATTTTCAATGAGGCGCATGATGATGATGATGATGATGATGATGATGAT 680
 QY 239 MetAlaSerTyrTrrpTrrpIleAsnLysAsnValLeu 250
 Db 681 GCCGCGATCTACTGCGGAGTAAACAAAGTTTGAATT 716
 RESULT 5
 H55010 928 bp mRNA linear EST 27-SEP-1999
 LOCUS H55010.1 Sorghum bicolor cv. TX430 Sorghum bicolor cDNA clone H55010
 DEFINITION 5' similar to adenine nucleotide translocator, chloroplast, mRNA
 sequence.
 ACCESSION H55010
 VERSION H55010.1 GI:1000790
 KEYWORDS EST.
 SOURCE Sorghum bicolor (sorghum)
 ORGANISM Sorghum bicolor
 BUKRYOTCA; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACMAD
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 928)
 MYRICH,R., DRESSEN,U., BROCKMANN,S., STREUBEL,M., CHANG,C., QIANG
 D., PATERSON,A.H. and WESTHOFF,P.
 The molecular basis of C4 photosynthesis in sorghum: isolation,
 characterization and RFLP mapping of mesophyll- and
 bundle-sheath-specific cDNAs obtained by differential screening
 Plant Mol. Biol. 37 (2), 319-335 (1998)
 JOURNAL MEDLINE PUBMED
 COMMENT
 Contact: Westhoff P
 Plant Molecular Biology
 Heinrich-Heine-University
 Universitaetsstrasse 1, D-40225 Duesseldorf, Germany
 Tel: (49)-211-311-2338
 Fax: (49)-211-311-4871
 Email: west@uni-duesseldorf.de
 Seq primer: SK (17-mer).

FEATURES
 source 1..928
 Location/Qualifiers
 /organism="Sorghum bicolor"
 /mol_type="mRNA"
 /strain="cv. TX430"
 /db_xref="taxon:4558"
 /clone="H55010"
 /clone_lib="Sorghum bicolor cv. TX430"
 /note="Vector: Lambda ZAP II; leaf cDNA library, unidirectionally cloned"
 BASE COUNT 225 a 182 c 245 g 275 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,52e-65 Length: 928
 Score: 694.50 Matches: 156
 Percent Similarity: 66.98% Conservative: 57
 Best Local Similarity: 49.06% Mismatches: 96
 Query Match: 26.41% Indels: 12
 DB: Gaps: 4
 US-09-869-433-2 (1-515) x H55010 (1-928)
 QY 187 LeuPheGlyIleGlyAlaAsnLysSerLeuLeuAlaSerGlyArgAlaIleValTrrpAla 206
 Db 3 CTGTTGGGCTTGGTGGCAATGTGCTTACATCTTCTGCGCGACGCGTAAATATTC 62
 QY 207 SerLysLeuArgAlaSerValSerGluGlyValAlaPProTTPGlyLysSerLeuArgLeu 226
 Db 63 TCAACATGAGAGAAATTTGGTCTCGAGTGTGGCAATTCGTTGAAG--- 119
 QY 227 LeuMetAlaMetThrIleValSerGlyLeuValLeuMetAlaSerTyrTrrpTrrpLeu 246
 Db 120 GCCATGACCATAGTGTGTACTGTGCTTATCATCACCGGTATCTATTTGGGCGGTGAC 179
 QY 247 LysAsnValLeuThrAspProArgPheTyrAsnProGluGluMetGluLeuGlyLysLys 266
 Db 180 AAGTTTGTATTGATCACTCATCTATC-----CCGTGCTCGACGGGAAAAAGA 227
 QY 267 GlyAlaLysProLysMetAsnMetLysAspSerPheLeuTyrLeuAspArgSerProTyr 286
 Db 228 AG-GACACAGCCAAAGCTCAGCATGGGTGAGACATGAAGTTCTGTCTCATCTCGTAT 286
 QY 287 IleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCysIleAsnLeuIleGluVal 306
 Db 287 GTGACGATCTTGGCCACATGTGCTGTCTTATGAAATTAACATTAACTTGTGAGAGTG 346
 QY 307 ThrTrrpLysSerGlnLeuLysLeuGlnTyrProAsnMetAsnAspTyrSerGluPheMet 326
 Db 347 ACATGGAATCAAAATTAAGACACAGCTT-CCAGCCCAATGAATATCTTCACTTCATG 405
 QY 327 GlyAsnPheSerPheTrrpThrGlyValValSerValLeuIleMetLeuPheValGly 346
 Db 406 GCGCATTTCTCAACGCGACCGCATAGCTTACCTTCAATGATGTTG---TTAGGAGA 462
 QY 347 AsnValIleArgLysPheGlyTrrpLeuThrGlyAlaLeuValThrProValMetValLeu 366
 Db 463 GTAATCCCAAAAGATTGGGTGGGAGTTGAGCTGAGTACACACCTGCGATGTACTT 522
 QY 367 LeuThrGlyIleValAlaPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuVal 386
 Db 523 CTCACAGAGATGGGTTTCTTCTCATGATTTTGTGTTGAGCCCTTACCTCTTATG 582
 QY 387 AlaMetPheGlyThrThrProLeuMetLeuAlaValAlaGlyAlaIleGlnAsnIle 406
 Db 583 ACCAAGTTTGGAGATGACCTTGTGCGGACAGCTTGTGTGGAGCAATCAGAACTT 642
 QY 407 LeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLysGluMetAlaTyrIlePro 426
 Db 643 TTCAGTAAAGATGCAAGATACAGCTGTTTATCTCTTGCACAAAGATGCGATACATTCCT 702
 QY 427 LeuAspGlnGluGlnValValLysGlyLysAlaIleAspValValAlaAlaArgPhe 446

Db 703 TTGGATGAGGATATGAAGGTTAAGGTAAGCAGTATTGACGTTGTCTGCAA-CGCCCTT 761

Qy 447 GILYSErSGLYGLYALALEUllEGlNGlNGlYLeuLeuValllEcySGLYserllEGly 466

Db 762 GGGAAATCGAGGCTGCTTGTGATCCAGCATTCATCTGCATTTGATCTCTGGG 821

Qy 467 AlAMEtPrOTYrLeuAlaValllELeuLeuPhelelleleAlallelrrLeuValser 486

Db 822 AATTGACACCCCTTACCTTGTGGAAATCTTGTGATATTGTTCTTGCAATGGCTGGGTGCT 881

Qy 487 AlArThrLYLeuAnLYSLeupheLeuAlaGlnSerAlaLeuLYSGlNGlNGlu 504

Db 882 GTAAGGTCCCTCGACTCACAGTTT-----TCAACCTGGCAAGCAAGAT 926

RESULT 6
AV926113 645 bp mRNA linear EST 18-JAN-2002
LOCUS AV926113 K. Sato unpublished cDNA library, cv. Haruna Nijo second
DEFINITION leaf stage seedling leaves Hordeum vulgare subsp. vulgare cDNA
clone based18c14 5', mRNA sequence.

ACCESSION AV926113
VERSION AV926113.1 GI:18221910
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
1 (bases 1 to 645)
Sato, K., Saitoh, D. and Takeda, K.
Barley EST sequencing project in NIG and Okayama Univ
Unpublished
Contact: Tadau Shin-ichi
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1..645
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Haruna Nijo"
/db_xref="taxon:112503"
/clone="based18c14"
/tissue_type="seedling leaves"
/dev_stage="second leaf stage"
/clone_1lb="K. Sato unpublished cDNA library, cv. Haruna
Nijo second leaf stage seedling leaves"

BASE COUNT 108 a 202 c 168 g 167 t

ORIGIN

Alignment Scores:
Pred. No.: 2,17e-64 Length: 645
Score: 684.00 Matches: 124
Percent Similarity: 77.57% Conservative: 42
Best Local Similarity: 57.94% Mismatches: 48
Query Match: 26.01% Indels: 0
Gaps: 0

US-09-869-433-2 (1-515) x AV926113 (1-645)

Qy 25 LysLYSVAlleuPrOMeThPheLeuWecPhePheCyslleThrPheAsnTYrThrValLeu 44

Db 2 AAGAAGATCGTGCCTCGGCTTCATGTCTCTGATCTCTTCAACTACACCATCTCG 61

Qy 45 ArgASPThrLYAspThrLeuLeuValGlyAlaProGlySerGlyAlaGluAlaLeuPro 64

Db 62 AGGGACACCAAGGCTGCTCGTCACCGCAAGGGAGCAGCGCCAGATCAATCCCT 121

Qy 65 PhelelLYSPhetrrLeuValAlProCysAlaIlellePheMetLeuIleTYrAlaLYs 84

Db 122 TTCCTCAAGACGTGGGTCACCTCCCATGGCCATCGGCTTCATCTCTACTCCAAAG 181

Qy 85 LeuSerAnlleLeuSerLYSGlAlaLeuPheTYrAlaValGlyThrProPheLeu 104

Db 182 CTCGGCAGCGCTCTCAAGAGGCGCTTTCACACCGTCATCTCCCTTCATCGCC 241

Qy 105 PhePheAlaLeuPheProThrYAlleTYrProLeuArgAspValLeuHisProThrglu 124

Db 242 TTCCTCGGCGCTTCGATACGTGCTACACCTTAAGCCGAGCCATCACCACCGGCG 301

Qy 125 PheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuGlyLeuValAlaIleLeu 144

Db 302 CTCGCCAGACGCGCTCGCTCGCTCGCCAGACTTCTGGGGCCCGCTGCCATCCTTA 361

Qy 145 ArgAsnTYrThrPheAlaAlaPheTYrValLeuAlaGluLeuTYrGlySerValMetLeu 164

Db 362 CGCGCTGAGATTCTTGCTCTTCACTTAATGGGGAGCTCTGGGGAGCGCTTCATC 421

Qy 165 SerLeuMetPheTYrGlyPheAlaAsnGluIleThrLYSlelleGluAlaLYsArgPhe 184

Db 422 TCCGTCCTGTTTGGGATTTGCCAATCAGATTACACGCTTGAAGAGGCTTAAAGATT 481

Qy 185 TYrAlaLeuPheGlylleGlyAlaAsnIleSerleuLeuAlaSerGlyArgAlaIleVal 204

Db 482 TACCACATGTTGGGCTTGGGGCCAAATGTGCTCATCTTCTGTCGACGCTGAAA 541

Qy 205 TrpAlaSerLYSLeuArgAlaSerValSerGluGlyValAspProTYrLYSerleu 224

Db 542 TACTTCTCAACATGAGCAGAAATTTGGGTCAGGGAGGAGCGAGTGGGCTTATTCATTG 601

Qy 225 ArgLeuMetAlaMetThrLYLeuAlaSerGlyLeuValLeu 238

Db 602 AAGGCATGATGACATAGTGTGTGTGCTGGGTTTCGTCATT 643

RESULT 7
B1177811 787 bp mRNA linear EST 07-MAR-2003
LOCUS B1177811
DEFINITION EST518756 cSTE Solanum tuberosum cDNA clone cSTE11L7 5' sequence,
mRNA sequence.

ACCESSION B1177811
VERSION B1177811.1 GI:14643622
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
1 (bases 1 to 787)
van der Hoeven, R., Bezzerides, J., Bachem, C., Visser, R., Cho, J.,
Chieningo, A., Bougri, O., Buell, C.R., Ronning, C., Tanksley, S. and
Baker, B.
Generation of ESTs from in vitro grown microtubers
Unpublished
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr. Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: M13F-R.
Location/Qualifiers

FEATURES
source
1..787
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Bintje"
/db_xref="taxon:4113"
/clone="cSTE11L7"
/tissue_type="axillary buds of stem explants; growing
sink-tubers"
/dev_stage="7, 8 and 10 days"
/lab_host="SOLR"
/clone_1lb="cSTE"

9 PheGlyLysLeuArgSerPheLeuTyrProIleHisThrHisGluLeuLysValLeu 28

Db 191 TGGGGTCGGGCTCGCCCAAGTTCCTGGGC-GTGAAGACGAGAGCGCTGAAGAGATCGG 249
Qy 29 PrometPheleuMetPhePheCysIleThrPheAsnTYRThrValLeuArgAspThrIys 48
Db 250 CGCGTGGGGCTCATCTCTTCTGCACTCCCTTCACTCAACCATCTCGGGGACGCCAAG 309
Qy 49 AspThrLeuIleValIGLYAlaIlePheSerGlyAlaGluAlaIleProPheIleIysPhe 68
Db 310 GACGCTGCTGCTGCTACCGCCCAAGGCGACAGCGGGAGATCATCTCCGTTCTCAAGAG 369
Qy 69 TrpLeuValIleProCysAlaIleIlePheMetLeuIleTYRAlaIleLeuSerAsnIle 88
Db 370 TGGGTCAACCTCCCATCGGCTTCATGCTGCTACACCAAGCTCGCCGACGTG 429
Qy 89 LeuSerIleGlnAlaLeuPheTYRAlaIleGlyThrProPheLeuIlePhePheAlaLeu 108
Db 430 CTCTCCAGAGAGCGCTCTTACACCGCTCATCTCCCTTCATCGCTTCTCGGCGCC 489
Qy 109 PheProThrValIleTYRProleuArgAspValLeuHisProThrGluPheAlaAspArg 128
Db 490 TTCGCTACCTGCTTACCCCATGCGCGACGCGCATCCACCGCGCGCTCGGACCGC 549
Qy 129 LeuGlnAlaIleLeuProGlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThr 148
Db 550 CTCCTCGCGCGGCTGGGCGCCAGCTTCCTGGGCGCCGCTCGCATCTCGGCTGAGAGC 609
Qy 149 PheAlaIlePheTYRValLeuAlaGluLeuTrpGlySerValMetLeuSerLeuMetPhe 168
Db 610 TTCTGCTCTTCTATGATGATGCGCGACGCTGGGCGACGCTGCTATCTCCGCTCTTC 669
Qy 169 TrpGlyPheAlaAsnGluIleThrIleHisGlnAlaIleAspArgPheTYRAlaLeuPhe 188
Db 670 TGGGATTCGCGCAATCACTTACTACGCTGAAGAGCGCAAGATTTTACCACCTGTC 729
Qy 189 GlyIleGlyAlaAsnIleSerLeuLeuAlaSerGlyAlaGlnAlaIleValTrpAlaSerIys 208
Db 730 GGGCTTGGGCGCAATGACACTCATCTCTCGGCTCGCAGATGAATTAATCTTCAAC 789
Qy 209 LeuAlaIleSerValSerGluGlyValAspProTrp 220
Db 790 ATGAGCGAGAAATTTGGTTCAGGCGGTGACGCGATG 825

RESULT 9
LOCUS CB604098 576 bp mRNA linear EST 16-MAY-2003
DEFINITION 3529_154_1_F08_Y_1 3529 - 2 mm ear tissue from Schmidt and Hake
ACCESSION CB604098
VERSION CB604098.1 GI:29543718
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoidae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 576)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY University
CONTACT Walbot V
JOURNAL Unpublished
COMMENT Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 3529_154_1_row: F column: 08.
LOCATION/Qualifiers
FEATURES
source
1..576
/organism="Zea mays"
/mol_type="mRNA"
/cultiVar="B73"

/db_xref="taxon:4577"
/tissue_type="ear"
/dev_stage="2 mm"
/lab_host="E. coli XLOLR"
/clone_id="3529 - 2 mm ear tissue from Schmidt and Hake
labg"
/note="Organ: ear; Vector: PAD-GAL4-2.1; Site 1: EcorI;
Site 2: XhoI; RNA isolated by Hake lab. 1 million pfu
amplified. Ampicillin is the selection marker."
BASE COUNT 101 a 199 c 138 g 138 t
ORIGIN
Alignment Scores:
Pred. No.: 2,03e-58 Length: 576
Score: 629.00 Matches: 114
Percent Similarity: 77.89% Conservative: 34
Best Local Similarity: 60.00% Mismatches: 42
Query Match: 23.92% Indels: 0
DB: 14 Gaps: 0
US-09-869-433-2 (1-515) x CB604098 (1-576)
Qy 21 ThrHisGluLeuIleValLeuProMetPheLeuMetPhePheCysIleThrPheAsn 40
Db 2 ACGAAGCAATCAAGAGATGCTGCCCTGGGGCTCATGTTCTTCTGCATCTCTCAAC 61
Qy 41 TYRThrValLeuAlaGAspThrIleValAspThrLeuIleValIGLYAlaIleProGlySerGlyAla 60
Db 62 TACACCAATCCCGCGGATACCAAGAGCGTGTCTGCTCAACGCCAAGGCGAGCGGCC 121
Qy 61 GlnAlaIleProPheIleIleTYRTrpLeuValAlaIleProCysAlaIleIlePheMetLeu 80
Db 122 GAGATCATCCGCTTCTCAAGACTGGGTCACTGCCCAAGGAGTTCATGCTC 181
Qy 81 IleTYRAlaIleSerAsnIleLeuSerIleGlnAlaLeuPheTYRAlaIleGlyThr 100
Db 182 CTCTACACCAAGCTCGCGCATGCTCTCCCAAGAGCGGCTTCTACGCGCTCATCTTC 241
Qy 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTYRProleuArgAspValLeu 120
Db 242 CGGTTCATCGCTTCTTGGGCGCTTGCCTACCTGCTCAACCCATGCGGAGCCCATC 301
Qy 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140
Db 302 CACCCCACTGGCTCGCGCATGCTCTCGCTCGCTCGGCGCCACCTTCTCGAGACC 361
Qy 141 ValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTYRValLeuAlaGluLeuTrpGly 160
Db 362 GTTCCCATCTCCCGCGCTGAGCTTGCCTTCTTACGATGAGCCGAGCTTGGGCGC 421
Qy 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrIleHisGly 180
Db 422 AGTGTCTCATATCCGCTTCTTGGGCGTTTCAATCAATGATTAATCACTTACAGTGAAGAG 481
Qy 181 AlaIleArgPheTYRAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
Db 482 GCCAAGAGTTCTACCGCGCTGTTGGGCTTGGGCGCATGAGCCCTCATCTTCTCGG 541
Qy 201 ArgAlaIleValTrpAlaSerIleValArg 210
Db 542 CGCACGGTGAATATTTCTCAACATGAGG 571

RESULT 10
LOCUS CD376868 893 bp mRNA linear EST 31-MAY-2003
DEFINITION PTM02037 Phaeodactylum tricornutum Uni-Zap XR Phaeodactylum
tricornutum cDNA 5', mRNA sequence.
ACCESSION CD376868
VERSION CD376868.1 GI:31252482
KEYWORDS EST.
SOURCE Phaeodactylum tricornutum
ORGANISM Phaeodactylum tricornutum
Eukaryota; stramenopiles; Bacillariophyta; Bacillariophyceae;

| REFERENCE | AUTHORS | TITLE | JOURNAL | MEDLINE | PUBMED | COMMENT |
|--|---|---|---|---------|---------|---------|
| 1 | Bacillariophycidae; Naviculales; Phaeodactylaceae; Phaeodactylum | 1 (bases 1 to 893) | | | | |
| | Scala, S., Carls, N., Falciatore, A., Chiusano, M.L. and Bowler, C. | Genome properties of the diatom Phaeodactylum tricornutum | Plant Physiol. 129 (3), 993-1002 (2002) | 2211123 | 1211455 | |
| | Contact: Bowler C | Laboratory of Molecular Plant Biology | | | | |
| | Stazione Zoologica "Anton Dohrn | | | | | |
| | Villa Comunale, I-80121, Napoli, Italy | | | | | |
| | Tel: 39 081 583 3268/3211 | | | | | |
| | Fax: 39 081 764 1355 | | | | | |
| | Email: chris@alpha.szn.it | | | | | |
| | Diatom EST Database (http://aves.chgen.szbwiler.com) | | | | | |
| | Seq primer: T3 backward | | | | | |
| | POLYA=yes. | | | | | |
| FEATURES | | | | | | |
| source | 1..893 | Location/Qualifiers | | | | |
| | /organism="Phaeodactylum tricornutum" | | | | | |
| | /mol_type="mRNA" | | | | | |
| | /db_xref="taxon:2850" | | | | | |
| | /cell_line="CCMP632" | | | | | |
| | /clone_lib="Phaeodactylum tricornutum Uni-Zap XR" | | | | | |
| | /note="Vector: Uni-Zap XR vector; Site_1: Eco RI; Site_2: | | | | | |
| | Xho I" | | | | | |
| BASE COUNT | 196 a 220 c 238 g 227 t | 12 others | | | | |
| ORIGIN | | | | | | |
| Alignment Scores: | | | | | | |
| Pidm. No.: | 1.49e-57 | Length: | 893 | | | |
| Score: | 623.50 | Matches: | 129 | | | |
| Percent Similarity: | 73.76% | Conservative: | 34 | | | |
| Best Local Similarity: | 58.37% | Mismatches: | 56 | | | |
| Query Match: | 23.71% | Indels: | 2 | | | |
| DB: | 14 | Gaps: | 1 | | | |
| US-09-869-433-2 (1-515) x CD376868 (1-893) | | | | | | |
| QY | 261 MetGlnTysGlyLysIleGluValThrTrpLysSerGlnLeuLysLeuGlnTyrProAsnMetLysAspSerPheLeuTyr | 280 | | | | |
| DB | 8 CTAAAGAACTATCCCTGACAAAAGAAAGCCCAAGTACGCTTAAGAAAGATGGCCAACTTT | 67 | | | | |
| QY | 281 LeuAspAspSerProTyrTrileuLeuLeuThrLeuLeuValIleAlaTyrGlyTleCys | 300 | | | | |
| DB | 68 CTTTTCAGACGCCCTTACATTCGCGATCTGCTTGTGTGTAATTGCTAGCGCATGTGC | 127 | | | | |
| QY | 301 IleAsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGlnTyrProAsnMetAsn | 320 | | | | |
| DB | 128 ATCAACATTGTCAATCGACTGAGAAAGCCAACTCAAGAGCGCTTCCCGATCCGAAAC | 187 | | | | |
| QY | 321 AspTyrSerGlnPheMetGlyAsnPheSerPheTrpThrGlyValValSerValLeuIle | 340 | | | | |
| DB | 188 TCGTACTCCGCCCTTTATGGTACTCTCTCGTACGTACCGG---GGGGCACCCTGATT | 244 | | | | |
| QY | 341 MetLeuPheValGlyGlyAsnValIleArgLysPheGlyTyrTrpLeuThrGlyAlaLeuVal | 360 | | | | |
| DB | 245 ATGATGTCTCGCGCGCATCCATCTTCAAGAAATTGGTTGGAGAAACAGCTGCGCTGTC | 304 | | | | |
| QY | 361 ThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn | 380 | | | | |
| DB | 305 ACTCCAAAGATGATCGGTGTGAGCGGGGTGGGTTTCTGCCCAACGATGTTTCGGGC | 364 | | | | |
| QY | 381 GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValAlaVal | 400 | | | | |
| DB | 365 GCCATCTCTCCGGTTCGCGCTCTTTTGGGACAAACCCCGCTTATGTTGGCGGTTCCTC | 422 | | | | |
| QY | 401 GlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLys | 420 | | | | |
| DB | 425 GGAGCGGCGCCAGAAATTTCTGACCAAGAGCTCCAAATATCTTTGTTGATCCGTGAAA | 488 | | | | |
| QY | 421 GluMetAlaTyrTrileProLeuAspGlnGlnGlnLysValLysGlyLysAlaAlaIleAsp | 440 | | | | |

| | | | | |
|------------------------|--|--|---|-----------------------------|
| Db | | 485 | GAAATGGCCCTTCATCCTTTGGATCAGGAATCCAAAACCAAGGCGGCATTGAT | 544 |
| Oy | | 441 | VatVala1Aa1AaAgPnehlYlvSergLY-GVAla1leu1llecInglngLYleuAuva | 460 |
| Db | | 545 | GTTGTGGCAATCCCTTGGGAAAAAGTGgggGCGTGTGATTAACAAGTTTATCTT | 604 |
| Oy | | 460 | lllecYglSerlleclglAlamectrPrOTyLeua1aValilleuPeuphaile1 | 480 |
| Db | | 605 | TGAGTCGTGTTGTTTAGCCGCCCTGATTCGCTATTTAGCCCTGATCTTGCTGTCTGCT | 664 |
| Oy | | 480 e 480 | : | |
| Db | | 665 A 665 | : | |
| RESULT 11 | | | | |
| LOCUS | | BM448586 | 685 bp | mRNA linear EST 01-APR-2002 |
| DEFINITION | | DSAO25G04.59793 An expressed sequence tag database for the halotolerant green alga, Dunaliella salina Dunaliella salina cDNA clone DSAO25G04 5, mRNA sequence. | | |
| ACCESSION | | BM448586 | | |
| VERSION | | BM448586.1 | GI:19854158 | |
| KEYWORDS | | EST. | | |
| SOURCE | | Dunaliella salina | | |
| ORGANISM | | Eukaryote; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Dunaliellaceae; Dunaliella. | | |
| REFERENCE | | 1 (bases 1 to 685) | | |
| AUTHORS | | Cushman,J.C. | | |
| TITLE | | An expressed sequence tag database for the halotolerant green alga, Dunaliella salina | | |
| JOURNAL | | Unpublished | | |
| COMMENT | | Contact: Cushman JC Department of Biochemistry University of Nevada MS200, Reno, NV 89557-0014, USA Tel: 775-784-1918 Fax: 775-784-1650 Email: jcushman@unr.edu PCR Primers FORWARD: T3 20mer BACKWARD: T7 21mer Seq primer: T3 20mer Seq primer: T3 20mer High quality sequence stop: 685. Location/Qualifiers | | |
| FEATURES | | source | | |
| | | 1..685 | /organism="Dunaliella salina" | |
| | | | /mol_type="mRNA" | |
| | | | /db_xref="taxon:3046" | |
| | | | /clone_id="DSAO25G04" | |
| | | | /tissue_type="Cells, which was adapted in 2.5M NaCl via a incremental series from 1.7 to 2.0 to 2.25 to 2.5 M NaCl, were exposed to 3.4 M NaCl for 5 hours" | |
| | | | /cell_type="Green" | |
| | | | /clone_lib="An expressed sequence tag database for the halotolerant green alga, Dunaliella salina" | |
| | | | /note="vector: Lambda Uni-Zap XR, Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; library construction was performed according to Stratagene's recommended protocol for the Lambda UniZapK vector and cDNA synthesis kit." | |
| BASE COUNT | | 123 a | 214 c | 164 t |
| ORIGIN | | | | |
| Alignment Scores: | | | | |
| Pred. No.: | | 3.67e-57 | | 685 |
| Score: | | 618.50 | | 123 |
| Percent Similarity: | | 73.81% | | Matches: 32 |
| Best Local Similarity: | | 58.57% | | Mismatches: 54 |
| Query Match: | | 23.52% | | Indels: 1 |
| DB: | | 12 | Gaps: | 1 |

QY 290 LeuThrLeuLeuValIleAlaIleTyrGlyIleCysIleAsnLeuIleGluValThrTrpIys 309
 Db 2 CTGGACAGCTGGTGGTGGCTAGCGTATTCATCACTCGTGGAGGTGACCTGGAG 61
 QY 310 SerGlnLeuLysLeuGlnTyrProAsnMetAsnAPRYSerGlnPheMetGlyAsnPhe 329
 Db 62 TCTAAGATCAAGGCCAGTGTCCCAACCCCAACGACTACTCTGTTCAATGGTAACTTC 121
 QY 330 SerPheTrpThrGlyValValSerValLeuIleMetLeuPheValGlyGlyAsnValIle 349
 Db 122 AGCAGACGACACTGGTGGACGTGACCTCGCCAGAGAGCTT---CTGTCCCGCTCATCTTC 178
 QY 350 ArgLysPheGlyTyrLeuThrGlyAlaLeuValThrProValMetValLeuThrGly 369
 Db 179 GCGAGATTTGGCTGGGGTACACTGCTCTCACTCACTCCACCTCTGCTGTCTCCCGT 238
 QY 370 IleValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPhe 389
 Db 239 CTGATTTTCTTCAGCTTGCTGATGTTCCCTAGGCCACATGCTCCATGCTGGCAGGCTGG 298
 QY 390 GlyThrTrpProLeuMetLeuAlaValAlaValGlyAlaIleGlnAsnIleLeuSerIys 409
 Db 299 GGCATGACCCCTGTATGCGCTGTGCTGGGCGCTGCCAGAACATCTTCAGCAAG 358
 QY 410 SerThrLysTyrAlaLeuPheAspSerThrLysGlnMetAlaTyrIleProLeuAspGln 429
 Db 359 TCTGCCAGATACAGCTGTTGACATCCCTGCCAAGAGATGGCTTAATCTCCCTTGATAT 418
 QY 430 GluGlnLysValLysGlyLysAlaIleAspValAlaAlaIleArgPheGlyLysSer 449
 Db 419 GAGGTCAAGACCAAGGTAAAGGCCCATTTGATGTGATCGACCCCTGGGGCAAGTCC 478
 QY 450 GlyGlyAlaLeuIleGlnGlnIleuLeuValIleCysGlySerIleGlyAlaMetThr 469
 Db 479 GGTGTGCTCGATCCACAGATTCATGATATGCTTGGCTCCCTGCGCCCTCCACA 538
 QY 470 ProTyrLeuAlaValIleLeuPheIleIleAlaIleTyrPheValSerAlaThrIys 489
 Db 539 CCTTACCTGGGCGCATCTGCTGGCCATTGTCGATCTGATTCACAGCCGCCCATCG 598
 QY 490 LeuAsnLysLeuPheLeuAlaGlnSerAla 499
 Db 599 CTGAACAAGCAGATTCGTGGCTCTCCAGCC 628
 RESULT 12
 BO813233 705 bp mRNA linear EST 01-AUG-2002
 LOCUS 1030034G10.y2 C. reinhardtii CC-1690, Deflagellation (normalized),
 DEFINITION Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
 BO813233
 ACCESSION BO813233.1 GI:22059043
 VERSION
 KEYWORDS
 SOURCE EST.
 ORGANISM Chlamydomonas reinhardtii
 Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadaceae; Chlamydomonas.
 1 (bases 1 to 705)
 REFERENCES
 Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre,
 P., McDermott, U.P., Strager, U., Sillow, D., and Stern, D.
 Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants. Project: 1030
 Unpublished
 JOURNAL Contact: Charles Hauser
 COMMENT DCMB Box 91000
 Duke University
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chauser@duke.edu
 FEATURES
 Source 1..705
 Location/Qualifiers

/organism="Chlamydomonas reinhardtii"
 /mol_type="mRNA"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_id="C. reinhardtii CC-1690, Deflagellation
 (normalized), Lambda Zap II"
 /note="Vector: pBluescript II SK-, Site 1: EcoRI, Site 2:
 XhoI; Deflagellation library, constructed by John Davies
 and Jeffrey McDermott, combines cDNAs from CC-1690 cells
 which had been re-synthesizing flagella for 15, 30 and 60
 min after being deflagellated by pH shock. PolyA mRNA was
 purified from each sample, pooled and cDNA synthesized.
 The cDNA was directionally cloned into Lambda Zap II.
 (Stratagene) in the EcoRI (5') and XhoI (3') sites.
 pBluescript II SK- plasmids were excised from the Lambda
 Zap clones by superinfection with Exsist (Stratagene)
 phage. The library was normalized using method 4 described
 in Bonaldo et al., (1996) Genome Research 6: 791-806."
 BASE COUNT 110 a 226 c 207 g 160 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4,92e-57 Length: 705
 Score: 617.50 Matches: 119
 Percent Similarity: 72.57% Conservative: 53
 Best Local Similarity: 50.21% Mismatches: 59
 Query Match: 23.48% Indels: 6
 DB: 13 Gaps: 3
 US-09-869-433-2 (1-515) x BO813233 (1-705)
 QY 141 ValAlaIleLeuArgAsnTyrThrPheAlaIlePheTyrValLeuAlaGluLeuTyrGly 160
 Db 11 GTTCCATTCGCGCACTGACCTTCTGCTGTCTACAGTGAAGGCGACCTGGGGC 70
 QY 161 SerValMetLeuSerLeuMetPheTyrGlyPheAlaIleGluIleTyrLysIleHisGlu 180
 Db 71 TCCGCTCGTGGTCCGCTGCTGTTCTGGGGCTTCCCAACGATCACACCCGCGAGAG 130
 QY 181 AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaIleSerLeuLeuAlaSerGly 200
 Db 131 GCTTCGAGTTCACCCCTGTTGGTGGGCGCCAAAGTGGCTGATTTCTCGGGC 190
 QY 201 ArgAlaIleValTyrPalaSerLysLeuArgAlaSerValSerGlnGlyValAspProTyr 220
 Db 191 CAGCGCGTGAAGTCTTCTGCGAGTCCGCGCGCTCGCCCGGATGTTGATGGCTGG 250
 QY 221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla 240
 Db 251 GGTGTGCCCTGAACGCGCTGATGGCATGTCCTCATCGCGCGCTGCTATTATCGGC 310
 QY 241 SerTyrTrpTyrPheLeuLysAsnValLeuThrAspProArgPheTyrAspProGluGlu 260
 Db 311 ATCTACTTCTGCTGCGAGCGCTGTGTGTC-----CCCAAGCTGAAGTCCGCGGAG 364
 QY 261 MetGlnLysGlyLysLysGlyAlaLysProLysMetAsnMetLysAspSerPheLeuTyr 280
 Db 365 GCGAAGAAGAAGAGAAC-----AAGACCAAGATGAGCGGCGAGTCTTGCGCTTC 418
 QY 281 LeuAspArgSerProTyrIleLeuLeuThrLeuLeuValIleAlaTyrGlyIleCys 300
 Db 419 CTGGCCAGTGCAGTCACTCCGCGACATGGCCACCTGTGGTGGCTTACGATATCGC 478
 QY 301 IleAsnLeuIleGlyValAlaThrTyrLysSerGlnLeuLysLeuGlnTyrProAsnMetAsn 320
 Db 479 ATCAACTGTGTGCGAGTGAACCTGGAAGGCAAGCTCAAGGCCCATCCCAACCCCAAC 538
 QY 321 AspTyrSerGlnPheMetGlyAsnPheSerPheTyrThrGlyValAlaSerValLeuIle 340
 Db 539 GACTACTGTCGTTCAATGGGGAATCTCCACCGCCACCGGTACCGGAGC---TTCACC 595
 QY 341 MetLeuPheValGlyGlyAsnValIleArgLysPheGlyTyrPheLeuThrGlyAlaLeuVal 360

Db 596 ATGATGATCTGTCCTCCGTCGATCTTCAAGCAGTTCGGCTGGNGTGTGCTGCCCTGAT 655

QY 360 1ThProValMetValLeuLeuThrgLylValAlphehAlaLeuVal 376

Db 656 CACCCCAATCAATGCTGATGATCAGGCTGCTGCTTCTTCACTCCGTGGTG 704

RESULT 13

CD423751 757 bp mRNA linear EST 02-JUN-2003

LOCUS SA1_1.E12.b1.A002 Salicylic acid-treated seedlings Sorghum bicolor

DEFINITION CDNA clone SA1_1.E12.A002 3', mRNA sequence.

ACCESSION CD423751

VERSION CD423751.1 GI:31330014

KEYWORDS EST.

SOURCE Sorghum bicolor (sorghum)

ORGANISM Sorghum bicolor

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Sorghum.

1 (bases 1 to 757)

Authors: Cordonnier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Salzman, R., Chua Tan, N., Gonzalez, M., Lane, S., Miller, V., Nanda, P., Olaseinde, O., Eastman, A. and Pratt, L.H.

An EST database from Sorghum: salicylic acid-treated seedlings unpublished

TITLE Contact: Cordonnier-Pratt MM

COMMENT Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-1271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmprratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.

Seq primer: Sug3 (CGACCTGACGCTCGACACAC)

POLYA=yes.

FEATURES

source

location/Qualifiers

1..757

/organism="Sorghum bicolor"

/mol_type="mRNA"

/cultivar="BTx623"

/db_xref="taxon:4558"

/clone="SA1_1.E12.A002"

/lab_host="DH10B-T1 phage-resistant E. coli"

/clone_lib="Salicylic acid-treated seedlings"

/note="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from seedlings grown in hydroponic culture. At 8 days of age, medium was supplemented with 1 mM salicylic acid (SA). Roots and shoots were harvested after 27 and 72 hr and material from both time points was combined prior to RNA isolation. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

BASE COUNT 191 a 155 c 191 g 220 t

ORIGIN

Alignment Scores:

Pred. NO.: 4.65e-56 Length: 757

Score: 609.00 Matches: 127

Percent Similarity: 69.28% Conservative: 40

Best Local Similarity: 52.70% Mismatches: 70

Query Match: 23.16% Indels: 4

DB: 14 Gaps: 2

US-09-869-433-2 (1-515) x CD423751 (1-757)

QY 265 LysIysGSLyAlAlYsProLyMetCasMetLysAspSerPheLeuTrLeuAspArgSer 284

Db 46 AAAAAGAGGACAAAGCCAAAGCTCAGCTGGAGACATGAAGCTTCTGTTCACCT 105

QY 285 ProTrIleLeuLeuThLeuThLeuValIleAlaTrYgYlLeCysIleAsnLeuIle 304

Db 106 CGGATGTGAGGAGATCTTCCACATGTCGTCTGCTTATGAAATAGCATTTACCTGTG 165

QY 305 GluValThrTrpLysSerGlnLeuLysLeuGlnTrpProAsnMetCasAspTrpSerGlu 324

Db 166 GAGGTGACATGGAATCAAAATGGAACACAGTCCCAAGCCCAAGAAATATCTTCA 225

QY 325 PheMetGlyAsnPheSerPheTrpThrGlyValValSerValLeuIleMetLeuPheVal 344

Db 226 TCACATGGCGATTTCTCAACTGCCACCGCATAGCTACTTTCATCATATGTTG--TTA 282

QY 345 GlyGlyAsnValIleArgLysPheGlyTrpLeuThrGlyAlaLeuValThrProValMet 364

Db 283 GCGAGAGTAATCTTCAGAAAGTTGGGGTGGAGAGTTGCAGCTACATCACCTCAGTG 342

QY 365 ValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGly 384

Db 343 TTACTTCTCACAGAGTGGGTTCTTCTCACTGATTTTGTGTGAGCCATTGACTCCT 402

QY 385 LeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValValAlaIleGln 404

Db 403 CTTATGACCAAGTTGGGATGACACCTTGGCTGGCGAGCTCATGTTGGAGCAATGCG 462

QY 405 AsnIleLeuSerLysSerThrLysTrpAlaLeuPheAspSerThrLysGlnMetAlaTrp 424

Db 463 AACATTTTCAGTAAGAGTCGCAAAAGTACAGCTGTGTTTATCTTCGCAAAAGATGCAATC 522

QY 425 IleProLeuAspGlnGlnGlnIleValLysGlyLysAlaAlaIleAspValValAlaAla 444

Db 523 ATTCCTTGGATGAGATGAAATGAAGTTAAAGCTTAAGACGATTTAGCTTGCAAC 582

QY 445 ArgPheGlyLysSerGlyValAlaLeuIleGlnGlnIleLeuValIleCysGlySer 464

Db 583 CCTTGGGGAATCTGGAGGTGCTTGTATCCAGCAGTTCATGATCTTCATTTGGACT 642

QY 465 IleGlyAlaMetCProTrpTrpLeuAlaValIleLeuLeuPheIleAlaIleTrpLeu 484

Db 643 CTTCGGAATTCGACACCTTACCTTGTGGAATTACTTCTGGATTTGTTCTTGATGGCTG 702

QY 485 ValSerAlaThrLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeuLysGlnGln 504

Db 703 GGTGCTGTAAGTCCCTGATCAAGTT-----TCAACCTGGCAAGCAAGAT 753

QY 505 Val 505

Db 754 CTC 756

RESULT 14

LOCUS AJ432835 678 bp mRNA linear EST 15-MAR-2002

DEFINITION AJ432835 S00007 Hordeum vulgare cDNA clone S0000700012A06F2, mRNA sequence.

ACCESSION AJ432835

VERSION AJ432835.1 GI:19521287

KEYWORDS EST.

SOURCE Hordeum vulgare

ORGANISM Hordeum vulgare

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; 1 (bases 1 to 678)

Authors: Saren, A.-M., Tanskanen, J., Paulin, L. and Schulman, A.H.

Title: Barley EST's

Journal: Unpublished

Comment: Contact: Schulman AH

Institute of Biotechnology

University of Helsinki
P.O. Box 56 (Valkinkari 6A), University of Helsinki FIN-00014,
Finland.

FEATURES

source Location/Qualifiers

1..678
/organism="Hordeum vulgare"
/mol_type="mRNA"
/db_xref="taxon:4513"
/clone="S000700012A06F2"
/dev_stage="Shoot"
/clone_1ib="S00007"
/note="2,-3,-4-days after germination"
BASE COUNT 168 a 139 c 166 g 205 t
ORIGIN

Alignment Scores:

Pred. No.: 5,15e-56 Length: 678
Score: 608.00 Matches: 116
Percent Similarity: 72.25% Conservative: 48
Best Local Similarity: 51.10% Mismatches: 59
Query Match: 23.12% Indels: 4
DB: 9 Gaps: 3

US-09-869-433-2 (1-515) x AJ432835 (1-678)

142 AAlaIleuAArgAsnTrpPhrheAlaAlaPhetYrValIleuAlaGluIleuTrpGlySer 161
142 AAlaIleuAArgAsnTrpPhrheAlaAlaPhetYrValIleuAlaGluIleuTrpGlySer 161
9 GGTATCCCTGAGATTGGAGCTTCTGCTTATTCATGTCATGCGAGCTATGGGCGAGC 68
162 ValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleuTrpGlySer 181
69 GTCGTTCTCTGCTCTCTCTCTGCGGTTTGGCAACCAATCACCACAGTTATGAGAGA 128
182 LysAspPheTrpAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGlyArg 201
129 AAAGAGTTCTACCTCTGTTGGCCCTTGGGCAAAATATGCCCCATTATCTTTCCGGGCGT 188
202 AlaIleValTrpAlaSerTrpLeuArgAlaSerValSerGluIleValAspProTrpGly 221
189 ACTGGAAGATTTCTCAAAATTTGGAGAACAGAACAGGCTCGACGAGTGTATGTTGGAG 248
222 IleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAlaSer 241
249 GTATCTTTGAAGAATATGATGGCATAGTACTTCTTGGCTTGTATCTTCTTCAT 308
242 TyrTrpTrpIleAsnLysAsnValIleuTrpAspProArgPheTrpAsnProGluMet 261
309 TATTGGGAGTGAACAGCTTGTGTTGAATGACCTTCTCTT-----CCAAATCTGAT 362
262 GlnLysGlyLysLysGlyAlaLysProLysMetAlaMetLysAspSerPheLeuTrpLeu 281
363 CACAAGAGAGAAAGAAC--AAACCTAAGCTTACATGAAGAGATGAAAGCTTCTG 419
282 AspArgSerProTrpIleLeuLeuLeuTrpLeuValIleAlaTrpGlyIleCysIle 301
420 ATCTCTTCAAAATATGATGGAGCCTTGTACTTATGTTGTCATACGCGATTAGTATC 479
302 AsnLeuIleGluValIleTrpLysSerGluLeuLysLeuGlnTrpProAsnMetAsnAsp 321
480 AATCTTGGAGAGTCAATGAAATCGAAGCTCAAGGACAGAGTCCCAAGTCCAAACAG 539
322 TyrSerGluPheMetGlyAsnPheSerPheTrpTrpGlyValIleValSerValLeuMet 341
540 TACTCATCTTCATGGGATTTCTCTACTGCACTGAGATTGCAACTTTCACGATGATG 599
342 LeuPheValGlyLysValIleArgLysPheGlyTrpLeuTrpGlyAlaLeuValThr 361
600 CTG---TTAGTCCGATTTATTTCCAAAGATTGGCTGGGCTGAGCCGCAATGATCAC 656
362 ProValMetValLeuLeuThr 368
657 CCCACAGTTTGTGTGACT 677

RESULT 15

BF460035

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES

source

1..744
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultiivar="Field grown Kuras"
/db_xref="taxon:4113"
/tissue_type="Tuber"
/clone_1ib="Mature tuber lambda ZAP"
/note="Vector: Lambda ZAP"
BASE COUNT 189 a 149 c 178 g 228 t
ORIGIN

Alignment Scores:

Pred. No.: 6.64e-56 Length: 744
Score: 607.50 Matches: 124
Percent Similarity: 66.80% Conservative: 47
Best Local Similarity: 48.44% Mismatches: 76
Query Match: 23.10% Indels: 10
DB: 10 Gaps: 2

US-09-869-433-2 (1-515) x BF460035 (1-744)

175 IleThrLysIleHisGluAlaLysArgPheTrpAlaLeuPheGlyIleGlyAlaAsnIle 194
3 ATCAAGACTGCGATGAGGCTTAAGAGATTATCTTGTGTTGAGCTTGAGCGCAATGTT 62
195 SerLeuLeuAlaSerGlyArgAlaIleValTrpAlaSerLysLeuArgAlaSerValSer 214
63 GCTCTTATTTCTCTGCGACAGTGAAGTCTTCTACCTTGAAGAGCTCTTTAGT 122
215 GluGlyValaAspProTrpGlyIleSerLeuArgLeuMetAlaMetThrIleValSer 234
123 CTGAGATTGATGGTGGCCATTTCTCTGAAGAAATGATGAGCATTTGTGATATG 182
235 GlyLeuValLeuMetAlaSerTrpTrpTrpIleAsnLysAsnValIleuTrpAspProArg 254
183 GGTGGGGAATCTGTTCTTACTGAGGAGTGAATGAATGTTGCTCCCAACTCGT 242
255 PheTrpAsnProGluGluMetGlnLysGlyLysLysGlyAlaLysProLysMetAsnMet 274
243 -----AGCAAGAGAGAGAGTAAACCTTAACATGACCA 278
275 LysAspSerPheLeuTrpLeuAspArgSerProTrpIleLeuLeuLeuVal 294

```
Db      279  ATGGAGAGACTTGAAGTCTTGATCTCTCAAAATATATCAGGATCTTGCCACATGTGTT 338
      295  ILeaLaTYrGlyIleCysIleasnleuileguValThrTrpYsSerGlnLeuLysLeu 314
      339  GTACCATATGGCATTAAGTATCAACCTTGTAAGTTACATGGAAGTCAAGCTCAAGCT 398
Qy      315  GlnTYrProAsnMetAsnAspTYrSerGlnPheMetGlyAsnPheserPheTrpThrGly 334
      399  CAGTCCCAAGCCCCAATGAATACTCTCATTCATGAGGAGACTTCTCACTGCTACTGGA 458
Qy      335  ValValSerValleuIleMetleuPheValGlyIleAsnValIleargLysPheGlyTrp 354
      459  ---ATAGCAACTTCATATATGATGTGTGTTAAGTCAATGATTTTCGACAGATGGGTGG 515
Qy      355  LeuThrGlyAlaLeuValThrProValMetValleuLeuThrGlyIleValPhePheAla 374
      516  GGACGACGAGCCCAAGATACACTACAGTCTTGCTCTTACCGGAGTTGTTCTTCTCC 575
Qy      375  LeuValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeu 394
      576  CTGCTTTTGTGTGGTGGCCCTCTAGCACCTACTCTTGCGAAGTTGGATGACTCTCTT 635
Qy      395  MetLeuAlaValIleValGlyAlaIleGlnAsnIleLeuSerIysSerThrLysTyrAla 414
      636  CTAGAGAGCTGTCTATGTGGTGCAATGCAGAACATTTC-AGTAAGAGTGCAAGATATAGT 694
Qy      415  LeuPheAspSerThrLysGluMetAlaTyrIleProLeuAspGlnGlu 430
      695  TTGTTTGACCCCTCGCAAGAAATGGGCTACATTCTTGTGATGAGGAC 742
Db
```

Search completed: November 25, 2003, 12:56:46
Job time : 3169 secs

THIS PAGE BLANK (USP 10)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 25, 2003, 12:10:00 / Search time 477 Seconds

(without alignments)
3557.160 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 2630

Sequence: 1 MTKTEKPKFLRSFLMPH.....AQSLKEQVQAEDSAPASS 515

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 219069 seqs, 1647345023 residues

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=xlh
-O=/cgn2_1/USPTO.epool/US09869433/runat_24112003_171501_18234/app_query.fasta_1.711
-DB=Published Applications NA -OPMT=fastcap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=BLOSUM62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTPMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09869433 @CGN_1_1_347 @runat_24112003_171501_18234
-NCPU=6 -ICPU=3 -NO MAP -LANG=ENGLISH -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------------------------|
| 1 | 2630 | 100.0 | 1637 | 9 | US-09-892-851-1 Sequence 1, Appl1 |

| | | | | | | |
|----|-------|------|---------|----|---------------------|--------------------|
| 2 | 1333 | 50.7 | 1875 | 10 | US-09-938-842A-694 | Sequence 694, App |
| 3 | 256 | 9.7 | 284 | 9 | US-09-294-093B-4236 | Sequence 4236, App |
| 4 | 242 | 9.2 | 246 | 9 | US-09-294-093B-5730 | Sequence 5730, App |
| 5 | 227.5 | 8.7 | 309 | 9 | US-09-294-093B-1312 | Sequence 1312, App |
| 6 | 162 | 6.2 | 1614 | 10 | US-09-738-626-117 | Sequence 117, App |
| 7 | 162 | 6.2 | 3309400 | 10 | US-09-738-626-117 | Sequence 1, Appl1 |
| 8 | 138.5 | 5.3 | 1389 | 9 | US-09-815-242-6064 | Sequence 6064, App |
| 9 | 138 | 4.84 | 484 | 11 | US-09-770-961-833 | Sequence 833, App |
| 10 | 133 | 5.1 | 1830121 | 14 | US-10-329-960-1 | Sequence 1, Appl1 |
| 11 | 132.5 | 5.0 | 1392 | 9 | US-09-815-242-6047 | Sequence 6047, App |
| 12 | 122.5 | 4.7 | 68750 | 13 | US-10-014-171-1 | Sequence 1, Appl1 |
| 13 | 120.5 | 4.6 | 1188 | 9 | US-09-815-242-9794 | Sequence 9794, App |
| 14 | 120.5 | 4.6 | 1278 | 9 | US-09-815-242-8073 | Sequence 8073, App |
| 15 | 120.5 | 4.6 | 76804 | 12 | US-10-085-959-109 | Sequence 109, App |
| 16 | 120 | 4.6 | 1503 | 10 | US-09-938-842A-1892 | Sequence 1892, App |
| 17 | 119.5 | 4.5 | 1434 | 9 | US-09-815-242-9742 | Sequence 9742, App |
| 18 | 119.5 | 4.5 | 5132 | 8 | US-08-781-986A-163 | Sequence 163, App |
| 19 | 119 | 4.5 | 1788 | 9 | US-09-815-242-4189 | Sequence 4189, App |
| 20 | 119 | 4.5 | 1815 | 9 | US-09-815-242-8427 | Sequence 8427, App |
| 21 | 119 | 4.5 | 12445 | 10 | US-09-070-927A-242 | Sequence 242, App |
| 22 | 118 | 4.5 | 1401 | 9 | US-09-815-242-9733 | Sequence 9733, App |
| 23 | 117.5 | 4.5 | 1290 | 9 | US-09-815-242-4623 | Sequence 4623, App |
| 24 | 117.5 | 4.5 | 3642 | 8 | US-08-781-986A-328 | Sequence 328, App |
| 25 | 117.5 | 4.5 | 77536 | 12 | US-09-940-316B-1 | Sequence 1, Appl1 |
| 26 | 116.5 | 4.4 | 1684 | 10 | US-09-881-752A-135 | Sequence 135, App |
| 27 | 116 | 4.4 | 1467 | 14 | US-10-156-761-4216 | Sequence 4216, App |
| 28 | 116 | 4.4 | 25801 | 12 | US-10-181-319-13 | Sequence 13, Appl1 |
| 29 | 116 | 4.4 | 9025608 | 14 | US-10-156-761-1 | Sequence 1, Appl1 |
| 30 | 115.5 | 4.4 | 1230 | 10 | US-09-738-626-361 | Sequence 361, App |
| 31 | 115 | 4.4 | 2124 | 10 | US-09-738-626-1524 | Sequence 1524, App |
| 32 | 115 | 4.4 | 3309400 | 10 | US-09-738-626-1524 | Sequence 1524, App |
| 33 | 115 | 4.4 | 12860 | 10 | US-09-070-927A-144 | Sequence 144, App |
| 34 | 114 | 4.3 | 1830121 | 14 | US-10-329-960-1 | Sequence 1, Appl1 |
| 35 | 113.5 | 4.3 | 3867 | 12 | US-10-171-319-10 | Sequence 10, Appl1 |
| 36 | 113 | 4.3 | 5668 | 9 | US-09-759-143-777 | Sequence 777, App |
| 37 | 113 | 4.3 | 5668 | 9 | US-09-780-669-777 | Sequence 777, App |
| 38 | 113 | 4.3 | 5668 | 9 | US-09-895-793-777 | Sequence 777, App |
| 39 | 113 | 4.3 | 5668 | 10 | US-09-895-814-777 | Sequence 777, App |
| 40 | 113 | 4.3 | 5668 | 12 | US-10-144-678A-777 | Sequence 777, App |
| 41 | 113 | 4.3 | 5668 | 12 | US-10-294-025-777 | Sequence 777, App |
| 42 | 113 | 4.3 | 5668 | 13 | US-10-012-896-777 | Sequence 777, App |
| 43 | 113 | 4.3 | 5668 | 14 | US-10-205-823-420 | Sequence 420, App |
| 44 | 113 | 4.3 | 5668 | 14 | US-10-205-823-420 | Sequence 420, App |
| 45 | 113 | 4.3 | 5668 | 14 | US-10-205-823-420 | Sequence 420, App |

ALIGNMENTS

RESULT 1
US-09-892-851-1
Sequence 1, Application US/09892851
Patent No. US20020081682A1
GENERAL INFORMATION:
APPLICANT: Aventis Pasteur Limited
TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
FILE REFERENCE: 72613-6
CURRENT APPLICATION NUMBER: US/09/892,851
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/114,060
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: US 60/123,967
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 60/141,271
PRIOR FILING DATE: 1999-06-30
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1637
TYPE: DNA
ORGANISM: Chlamydia pneumoniae
FEATURE:
NAME/KEY: CDS
LOCATION: (51)..(1595)

US-09-892-851-1

Alignment Scores:
 Pred. No.: 9,04e-294 Length: 1637
 Score: 2630.00 Matches: 515
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-09-869-433-2 (1-515) x US-09-892-851-1 (1-1637)

QY 1 MetThrLysThrGluGluLysProPheGlyLysLeuArgSerPheLeuTrpProIleHis 20
 DB 51 ATGACAAAACCGAAGAAAACCTTTTGGAAAATTGCGCTCTTCTTGCGCGCATACAT 110
 QY 21 ThrHisGluLeuLysValLeuProMetPheLeuMetPhePheCysIleThrPheAsn 40
 DB 111 ACTCAGAGCTAAGAAAGATTCTGCCAATGTTCTTAATGTTCTTCTGTATTAATTAAAC 170
 QY 41 TyrThrValLeuArgAspThrLysAspThrLeuIleValGlyAlaProGlySerGlyAla 60
 DB 171 TATACGGTGTACGGGATCAAAAGACACTTATGTGGGAGCTCTGGTCTGTGCA 230
 QY 61 GluAlaIleProPheIleLysPheTrpLeuValValProCysAlaIleIlePheMetLeu 80
 DB 231 GAGGCAATACCTTTCAATCAAGTTTGCTGTGTTGCCCTGTGCTATTAATGCTT 290
 QY 81 IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr 100
 DB 291 ATTATGCAAGACTAAGTAATATTTTAAGTAAGCAGGCTTATTTATGCAAGTGAACG 350
 QY 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
 DB 351 CCTTTTAAATTTCTTGTCCCTGTCCGAGTGAATTAATTCCTCCAGCGCATGTCTTAA 410
 QY 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProGlyLeuLeuGlyLeu 140
 DB 411 CATCTTACGAATTTGCTGACCGCTTTACAGGCCATCTTACCTCCAGGATGCTAGACATC 470
 QY 141 ValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTyrValLeuAlaGluLeuTrpGly 160
 DB 471 GTTGCATCTTAAGAAAGCTGACATTTGCTGATTTAATGTAATGCTGCAATAGGGGA 530
 QY 161 SerValIleLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleHisGlu 180
 DB 531 AGCGTCAGCTATCTCTATATGTTGCGGATTTGCTAATGAATAACAAAATCCACGAA 590
 QY 181 AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
 DB 591 GCAGAGCGTTCTTACGCTCTTTCCGATCGAGCTAATATTTCTTTACTACCTTCTGCT 650
 QY 201 ArgAlaIleValTrpAlaSerLysLeuArgAlaSerValSerGlyValAlaAspTrp 220
 DB 651 CGTGAATGTTGGGCTTCAAAAGTTGAGAGCTTCCGTTTCAAGAGGTAGATCTCTGG 710
 QY 221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla 240
 DB 711 GGAATTTCTTAACGCTTTTGAATGACTATGATGAGCTTATGAGCTTCTTATGAGCC 770
 QY 241 SerTyrTrpTrpIleAsnLysAsnValLeuThrAspProArgPheTyrAsnProGlyGlu 260
 DB 771 AGTTACTGGTGATCAATAGAACGTATTTGACCGATCTCTCGTTCTATATATCAGAAAGA 830
 QY 261 MetGlnLysGlyLysLeuGlyAlaLysProLysMetAsnMetLysAspSerPheLeuTyr 280
 DB 831 ATGCAAAAAGGGAAGAAAGGTGCTAAACCTAAATGAATATGAAGATAGCTTCTCTAT 890
 QY 281 LeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCys 300
 DB 891 CTGTATAGATCTCTTATATATTTTAACTCTCTGTTATATGCTATAGTATTTGCT 950
 QY 301 IleAsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGlnTyrProAsnMetAsn 320

DB 951 ATTAACCTTAATCGAAGTACTGGAAAAGTCAGCGTAAACGCAATATCTTAATGTGAT 1010
 QY 321 AspTyrSerGluPheMetGlyAsnPheSerPheTrpThrGlyValValSerValLeuIle 340
 DB 1011 GACTATATGTGATTCATGGGAACTTCTCTTCTGAGTCGGCGTAGATCCGATCTTATTC 1070
 QY 341 MetLeuPheValGlyAlaAsnValIleArgLysPheGlyTyrTrpLeuThrGlyAlaLeuVal 360
 DB 1071 ATGCTATTGTGGGTGAGTCACTTGTAATTTGGATGGTTACTGAGCCCTAGTC 1130
 QY 361 ThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn 380
 DB 1131 ACTCTGTCAATGATCTCTCTTAACAGATATCGTTTCTTCCGCTCTGTTATCTTAAGAAC 1190
 QY 381 GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValVal 400
 DB 1191 CAAGCTTCTGGCGTGTGCTATGTTCGTACAACTCCTCAAGCTAGCTGTGGTGTTC 1250
 QY 401 GlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLys 420
 DB 1251 GAGCTATACGAATATTTCTTCAAAATCCAAATAATGCTCTCTTTGACTCAACTTAA 1310
 QY 421 GluMetAlaTyrIleProLeuAspGlnGluLysValLysGlyLysAlaAlaIleAsp 440
 DB 1311 GAAATGGCTATATCCCTCTTGAACCAAGCAAGCAAAAGTAAAGGTAGGCTGATTCAT 1370
 QY 441 ValValAlaAlaArgPheGlyLysSerGlyGlyAlaLeuIleGlnGlnGlyLeuLeuVal 460
 DB 1371 GTAGTTGCCCGCCCTTCGAAAAATCAGAGAGACTTAAATCCAAAGTTTCTGCTCT 1430
 QY 461 IleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValIleLeuLeuPheIleIle 480
 DB 1431 ATCTGGAAATATTTGAGCTATGACCCCTTAATCTTGCAGATATCTTCTTATCATCTT 1490
 QY 481 AlaIleTrpLeuValSerAlaThrLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeu 500
 DB 1491 GCTATTTGGTTGGTTTTCGCAACTAAGTTAAACAAATATCTTAAGCCAGCTGCTCTT 1550
 QY 501 LysGluGlnGluValAlaGlnGluAspSerAlaProAlaSerSer 515
 DB 1551 AAAAGACAAAGAAAGTGGCTCAAGAAAGATTCACTCTCTCTCTCA 1595

RESULT 2
 US-09-938-842A-694
 ; Sequence 694, Application US/0993842A
 ; Patent No. US20020160378A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Harper, Jeff
 ; APPLICANT: Krebs, Joel
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 ; TITLE OF INVENTION: SAME, AND METHODS OF USE
 ; FILE REFERENCE: SCRP.300-3
 ; CURRENT APPLICATION NUMBER: US/09/938, 842A
 ; CURRENT FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 60/227, 866
 ; PRIOR FILING DATE: 2000-08-24
 ; PRIOR APPLICATION NUMBER: US 60/264, 647
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/300, 111
 ; PRIOR FILING DATE: 2001-06-22
 ; NUMBER OF SEQ ID NOS: 5379
 ; SEQ ID NO 694
 ; LENGTH: 1875
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 US-09-938-842A-694

Alignment Scores:
 Pred. No.: 1.32e-143 Length: 1875
 Score: 1333.00 Matches: 259

Percent Similarity: 69.60% Conservative: 89
 Best Local Similarity: 51.80% Mismatches: 140
 Query Match: 50.68% Indels: 12
 DB: 10 Gaps: 3

US-09-869-433-2 (1-515) x US-09-938-842A-694 (1-1875)

QY 16 LeuTriProteinH1STH1SLGLLeuLysValLeuProMetPheLeuMetPhePhe 35
 DB 313 ATTTCGGTGTGAGGTTGCAACCTTGAAAGAAATATCCCTTTAGAGATGATGTTCTT 372
 QY 36 Cys1LeuThrPheAsnYrThrValLeuArgSerPheThyAspThrLeuIleValAla 55
 DB 373 TGTATTCCTTTCAATACACATTCCTGAGGATACAAAGATGCTTGTTGCTGACGCG 432
 QY 56 ProGlySerGlyAlaGlyAlaIleProPheIleLysPheThrLeuValProCysAla 75
 DB 433 AAAGGAAGTTCTCGTGAATTAATACCTTTCTGAAAGACTTGCGAATCTTCTATGCGC 492
 QY 76 IleIlePheMetLeuIleYrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPhe 95
 DB 493 ATTGGGTTATGCTCTCTACACTTAACCTTCCAAATGTTCTCTCAAGAGGCTCTGTT 552
 QY 96 TyrAlaValGlyThrProPheLeuIlePhePheAlaLeuPheProThrValIleTyrPro 115
 DB 553 TACACTGTATATGCTCTTTCATCATCTACTTGGGCTTGGTTGCTGATGATACCT 612
 QY 116 LeuArgAspValLeuHisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProPro 135
 DB 613 CTCGACACTATATTCACCCGGAAGCTCTCCAGATTAAGCTCTTACCAACCTCGGCCCA 672
 QY 136 GlyLeuLeuGlyLeuValAlaIleLeuArgSerThrPheAlaAlaPheTyrValLeu 155
 DB 673 AGATTACGGGCTCTATTCGAAATATGCGGATTTGGGATTTCTGTTTATGATATG 732
 QY 156 AlaGluLeuTriProGlySerValMetLeuSerLeuMetPheTriProGlyPheAlaAsnGlnIle 175
 DB 733 GCTGAGCTTTGGGGTATGCTGTGCTGCTCACTTCTTGGGCTTGGCTTCAATCAAGATC 792
 QY 176 ThrLysIleHisGlyAlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSer 195
 DB 793 ACAACTGTGATGAGCCAGCAAAATTCATCTTGTGCGGACTTGAGCGCAATGTTGCA 852
 QY 196 LeuLeuAlaSerGlyArgAlaIleValIlePheAlaSerLysLeuArgAlaSerValSerGlu 215
 DB 853 CTGATTTTCTCAGGAAGAACCGTGAATATCTTCTTAACCTTGAAGAAATCTTGCTCT 912
 QY 216 GlyValAspProTriProGlyLysSerLeuArgLeuMetAlaMetThrIleValSerGly 235
 DB 913 GGAATTACCGCTGGGACAGTTCTTGAAGCCATGATGACATTTGGTGGGAATGGGA 972
 QY 236 LeuValLeuMetAlaSerYrTriProIleAsnLysAsnValLeuThrAspProArgPhe 255
 DB 973 CTCGCCATTTGTCCTCTATTTGCTGAGGTCAATGATATGTTCTCTTCCAAACCGGATC 1032
 QY 256 TyrAsnProGluGluMetGlnLysGlyLysValAlaLysProLysMetAsnMetLys 275
 DB 1033 AAGAAACAGAGAGG-----AAACCGAAGATGGAACAGATG 1068
 QY 276 AspSerPheLeuYrLeuAspArgSerProTyrIleLeuLeuLeuThrLeuValIle 295
 DB 1069 GAAAGCTTGAAGTTCTTGATATCATCACATACATTAGAGATCTTGCTACTTTAGTGTG 1128
 QY 296 AlaTyrGlyIleCysIleAsnLeuIleGluValThrTriProLysSerGlnLeuValLeuGln 315
 DB 1129 GCATACCGATTAATATCAATCTTGGAAGTCACTGAAGAAATCAAGACTTAAGCTCAG 1188
 QY 316 TyrProMetAsnAspYrSerGluPheMetGlyAsnProPheTriProThrGlyVal 335
 DB 1189 TTCCCTAGCCCGAATGATCTCAGCATTTATGAGAGCTTCTCAACCTGACGGGTT 1248
 QY 336 ValSerValLeuIleMetLeuPheValGlyLysAsnValIleArgLysPheGlyTriProLeu 355

DB 1249 GCAACATTCAACATGATGCTT---CTCAGCAATACGATTAATCAATAGTATGTTGGGA 1305
 QY 356 ThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleValIlePhePheAlaLeu 375
 DB 1306 GTAGCTGCAAGATACACCCCAACCTGCTGATGATGCTGATGCTGCTTCTCTCTA 1365
 QY 376 ValIlePheArgAsnGlnIleAsnSerGlyLeuValAlaMetPheGlyThrProLeuMet 395
 DB 1366 ATATGTTTGGGGCCCATTCGACCACTTGTGCCAAGCTGTGATAGACACCGCTACTT 1425
 QY 396 LeuAlaValValGlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyrAlaLeu 415
 DB 1426 GCAGCTGTATGTCGGTGCCTTCAGATATCTTCAAGAAAGTGCACAGTCAAGCTTG 1485
 QY 416 PheAspSerThrLysGluMetAlaTyrIleProLeuAspGlnGluGlnValLysGly 435
 DB 1486 TTGACCTTGCAGAAAGAAAGGCTTAATCCATGATGATAGACACCAAGTTAAAGGC 1545
 QY 436 LysAlaAlaIleAspValValAlaAlaArgPheGlyLysSerGlyValAlaLeuIleGln 455
 DB 1546 AAAGCTGCAATGAGTGTGCTGCAACCCATTAGGGAATCAGGGAGCTTATATACG 1605
 QY 456 GlnGlyLeuLeuValIleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValIle 475
 DB 1606 CAGTTCAATGATCTTATCTTGTGATGATCTAGCAATTCACGCCGATCTTAGAATATG 1665
 QY 476 LeuLeuPheIleIleAlaIleTriProValSerAlaThrLysLeuAsnLysLeuPheLeu 495
 DB 1666 TTGTTGTTATTTGTCACCTGCTGCTGATGCTGACAGCTAAGTCGCTGGAGGACAGTTC 1722
 QY 496 AlaGlnSerAlaLeuLysGlnGluGlnValAlaGlnGlnLysSerAlaProAlaSerSer 515
 DB 1723 -----AACGCTTGGCTCTGAAAGAGACTTGAGAGGAATGAGAGACTTCTACG 1776

RESULT 3
 US-09-294-093B-4236
 Sequence 4236, Application US/09294093B
 Patent No. US2001005135A1
 GENERAL INFORMATION:
 APPLICANT: LalGudi, Raghunath, V.
 APPLICANT: Ico, Laura, Y.
 APPLICANT: Sherman, Bradley, K.
 TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
 FILE REFERENCE: PL-0009 US: US/09/294,093B
 CURRENT APPLICATION NUMBER: US/09/294,093B
 PRIOR FILING DATE: 1999-04-16
 PRIOR APPLICATION NUMBER: 60/082,567
 NUMBER OF SEQ ID NOS: 6207
 SOFTWARE: PERL Program
 SEQ ID NO 4236
 LENGTH: 284
 TYPE: DNA
 ORGANISM: Zea mays
 FEATURE:
 NAME/KEY: misc_feature
 OTHER INFORMATION: Incyte ID No. US2001005135A1 700353546H1
 LOCATION: 132, 274
 OTHER INFORMATION: a, t, c, g, or other
 US-09-294-093B-4236

Alignment Scores:
 Pred. No.: 284
 Score: 256.00
 Percent Similarity: 72.34%
 Best Local Similarity: 59.57%
 Query Match: 9.73%
 DB: 9 Gaps: 0

US-09-869-433-2 (1-515) x US-09-294-093B-4236 (1-284)
 QY 377 IlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeu 396

```
Db 3 TTGTTGGTGAGCCATTGACCTCTTATAGCAAGTTGGGATACGCTTTGCTTGGC 62
Qy 397 AAlaValValGlyAlaIleGlnAsnIleLeuSerThrIleTyrAlaLeuPhe 416
Db 63 GCAGCTTATGTTGGAGCATGAGCAATTTTCTACATAAGAGCAAAATATACAGTCTGTTT 122
Qy 417 AapSerThrIleGlnMetAlaTyrIleProLeuAspGlnIleGlnIleValIleGlyIle 436
Db 123 GATCTTCGCAAGATGCGATGCTTCTTGTGATGAGATATGAGGTGAAAGTAAA 182
Qy 437 AAlaIleAspValValAlaAlaArgPheGlyIleSerGlyIleAlaIleGlnIle 456
Db 183 GCGGCTATGATGTTGTGTGCAACCCCTGGGGAATCTGGAGGTGCTGTATCCAGCAG 242
Qy 457 GlyLeuLeuValIleCysGlySerIleGlyAlaMetThrPro 470
Db 243 TTCATGATCCCT-GTCATCGGTTCTCTCGCGAANTGACACCG 283

RESULT 4
US-09-294-093B-5730
; Sequence 5730, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Laljundi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 5730
; LENGTH: 246
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700382240H1
US-09-294-093B-5730

Alignment Scores:
Pred. No.: 9.34e-19 Length: 246
Score: 242.00 Matches: 45
Percent Similarity: 74.68% Conservative: 14
Best Local Similarity: 56.96% Mismatches: 20
Query Match: 9.20% Indels: 0
Gaps: 0

US-09-869-433-2 (1-515) x US-09-294-093B-5730 (1-246)
Qy 348 ValIleArgIlePheGlyIleThrIleValLeuValThrProValMetValLeu 367
Db 6 ATCCTCAGAAAGTTGGGTGGGAGCTTGCACGTACATCAAGCCCTGCACTTACTCTC 65
Qy 368 ThrGlyIleValPhePheAlaLeuValIlePheArgAsnIleAlaSerGlyLeuValAla 387
Db 66 ACAAGAGTTGGGCTTCTCTCACTGATTTGTTGGTGAAGCATTCGCTCTTATGACC 125
Qy 388 MetPheGlyThrThrProLeuMetLeuAlaValValIleGlyAlaIleGlnAsnIleLeu 407
Db 126 AAGTTTGGGATGACACCTTCTTCCGCGCAGTCTATGTTGAGCAATGCAAGATTTTC 185
Qy 408 SerIleSerThrIleTyrAlaLeuPheAspSerThrIleGlnMetAlaTyrIlePro 426
Db 186 AGTAAAGTGCAGAAATATACAGTCTGTTGATCTCTTGCAAGAGATGCGATTAATTCCT 242

RESULT 5
US-09-294-093B-1312
; Sequence 1312, Application US/09294093B
```

```
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Laljundi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 1312
; LENGTH: 309
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700344107H1
US-09-294-093B-1312

Alignment Scores:
Pred. No.: 6.36e-17 Length: 309
Score: 227.50 Matches: 46
Percent Similarity: 66.67% Conservative: 24
Best Local Similarity: 43.81% Mismatches: 32
Query Match: 8.65% Indels: 3
Gaps: 2

US-09-869-433-2 (1-515) x US-09-294-093B-1312 (1-309)
Qy 183 ArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGlyArgAla 202
Db 1 AAATTTACCCGCTATTGCTGGGCTTAATTTCTCTTATCTTTTCTGGGCTACT 60
Qy 203 IleValTyrPalaSerIleValArgAlaSerValSerGlyIleValAlaAspProTyrGlyIle 222
Db 61 GTGAAGTATTTCTCAATTTGCGCAAGACATTTGGCTCTGGAATGATGCTGGGAAGTA 120
Qy 223 SerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValIleMetAlaSerTyr 242
Db 121 TCTTTGAAGAAATGATGACATGATGCTCTCTTGGACTTGTCAATACCTTCATCTAT 180
Qy 243 TrpTyrIleLeuIleValAsnValIleThrAspArgPheTyrAsnProGlnIleMetGln 262
Db 181 TGGGAGATGAAACAAGTTGTTGATGATGATCTTCACTT-----CCAAAGGCTGATGT 234
Qy 263 LysGlyIleValIleGlyAlaIleValProLysMetCAsnMetLysAspSerPheLeuTyrLeuAsp 282
Db 235 AAG---AAGAAAAAGAAAAGCCCTAAACTTGCGATGAAGAAGACTCGAAGTTCTGCTC 291
Qy 283 ArgSerProTyrIle 287
Db 292 TCTCGAGGTATGTG 306

RESULT 6
US-09-738-626-117
; Sequence 117, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
```



```

; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 117
; LENGTH: 1614
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-117

Alignment Scores:
Pred. No.: 3,24e-08 Length: 1614
Score: 162.00 Matches: 127
Percent Similarity: 34.88% Conservative: 83
Best Local Similarity: 21.10% Mismatches: 210
Query Match: 6.16% Indels: 182
DB: 10 Gaps: 29

US-09-869-433-2 (1-515) x US-09-738-626-117 (1-1614)

Qy 2 ThrlystrHrIgluLys-----ProPheGlyLysLeuArg----- 13
Db 16 TCMAAAACCTTCAAGAAACCAACAGCCAGACGACCGACGACGACCTAGTAAATGAGC 75
Qy 14 ---SerPheLeuTrpIleHisThr-----HisGluLeuLysValLeuPromet 30
Db 76 GATAAATTTTCATAGCGTGGAACACTTGAAGAAACAGCTCCACACCACTTACGCTT 135
Qy 31 PheLeuMetPhePheCysIleThrPheAsnTrpThrValLeuAspThrLysAspThr 50
Db 136 TTTCTTTACTTTCTTCGATCAGCGCGCTGCTTCATCATGAGCGTGATGAACGTA 195
Qy 51 LeuIleValGlyAlaProGlySer----- 58
Db 196 TCCGCGATT---GTTCCAGGCTCTGATGAGAACTATTGTAAAGCGCTGTCAACCGT 252
Qy 59 -----GlyAlaGluAlaIleProPheIleLysPhe 68
Db 253 GAAGGCTTACCTGTTGACTACGAATCTCGGGCAACCTACATCGGTTCCCGCGCTG 312
Qy 69 TrpLeuValValProCysAlaIle-----IlePheMet 79
Db 313 CTCACCGGTGTCGAATTTCTGTGCAAGTGGTGTCTGAACGTTCCGCGATGTGGCT 372
Qy 80 LeuIleTrpAlaLysLeuSerAsnIleLeuSerLysGluAlaLeuPheTrpAlaValGly 99
Db 373 GCGCTGATTAAGAAACTTTTGTTCGGCGAAAGATCGTTTGCATTAAGCAATGCGGT 432
Qy 100 -----ThrProPheLeuIlePhePheAlaLeuPheProThrValIleTrpPro 115
Db 433 GTGATTGGCGTGAACCGGCTGATCATGCGGACCGCTGCTTC---GTGATGTCACCACT 489
Qy 116 LeuArgAspValLeu-----HisProThrGluPheAlaAspArgLeu 129
Db 490 TTGGCCCGCATGCGTTTAAAGCTGCTGTCGCGACCCCTGTG----- 531
Qy 130 GluAlaIleLeuProPheGlyLeuGlyLeuValAlaIleLeuArgAsnTrpThrPhe 149
Db 532 -----GCTGGGCTAATGGTTCGTTTCGACGCTGTG----- 561
Qy 150 AlaIlePheTrpValLeuAlaGluLeuTrpGlySerValMetLeuSerLeuMetPheTrp 169
Db 562 GGTTCAGATTAATTCACAGCATTTGTG-----CCACACCACTTGTGACCTTTTTCG 615
Qy 170 GlyPheAlaAsnGluIle-----ThrLysIleHisGlu 180
Db 616 GGAATTACCAACGCGGTGATGAGACACTTCCAGGCATTTGCGACAACTGAAGTCAATCCG 675

```

```

Qy 181 AlaLysArgPheTrpAlaLeuPheGlyTrpIleGlyAlaAsnIleSerLeu---LeuAlaSer 199
Db 676 GTTTCATCACTATATC-----TTCAATATTCATCTCGATGTGTGGCTGTGTATGT 729
Qy 200 GlyArgAlaIle-----ValTrpAlaSerLysLeuArgAlaSer 212
Db 730 GGTTCCTCATTAATAGAGTCTGTGAACCTCGGATGTGGCTCAGAAAATGCTTACGAG 789
Qy 213 ValSerGluGlyValAspProTrpGly-----IleSerLeuArgLeu 226
Db 790 TATCAGAAAGCATTTGAACCCACGACGACAGATGATGAGAAATCTCTGCAACCTTA 849
Qy 227 -----LeuMetAlaMetThr 231
Db 850 ACCGACAGAGAAACCGCGCGCTGACATTTTCATGTGACACCTCGGACGCGCATC 909
Qy 232 IleValSerGlyLeuValIleMetAlaSerTrpTrpIleAsnLysAsnValLeuThr 251
Db 910 ATCCTGCTGCTGTGGTGTCTGATTCGCGGATCCCATGAGAAATGAGAT----- 960
Qy 252 AspProArgPheTrpAsnProGluGluMetGlnLysGlyLysGlyAlaLysProLys 271
Db 961 -----GGTGAATCTTGCCTACC 978
Qy 272 MetAsnMetLysAspSerPheLeuTrpLeuAspArgSerProTrpIleLeuLeuThr 291
Db 979 TCGCCACTGCTGACCTCTGTGTGTATTT-----GTAATTTGTGTTTTCATG 1026
Qy 292 LeuLeuValIleAlaTrpGlyIleCysIleAsnLeuIleGluValThrTrpLysSerGln 311
Db 1027 GTGATGGCGCTGCGCTACCGCATGTGTGTGGACGATCAAG----- 1068
Qy 312 LeuLysLeuGlnTrpProAsnMetAsnAspTrpSerGluPheMetGlyAsnPheSerPhe 331
Db 1069 -----AACATGATGATGTCGTGAACATGATGAGGCGAAGCAATC--- 1107
Qy 332 TrpThrGlyValValSerValLeuIleMetLeuPheValGlyValAsnValIleArgLys 351
Db 1108 ---AAGACATGATGTGTTCTTGTGTGTGGCTTCATTTTGGACAGTTGTGGCGCTG 1164
Qy 352 PheGlyTrpLeuTrpGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleVal 371
Db 1165 TTTTAACTCG-----ACGGGCAATGCT 1185
Qy 372 PhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeu-----ValAlaMetPhe 389
Db 1186 ACCTGACCTGCTGT-----CAGGTCCTCGGATTTGGAAGCATCGGCTTACC 1236
Qy 390 GlyThrThrProLeuMetLeuAlaValAlaValAlaIleGlnAsn-----IleLeu 407
Db 1237 GGAATTCCTCGCATGATGATTAATTAATTTGCGCATGTTGAACGCTGCTGATTAAT 1296
Qy 408 SerLysSerThrLysTrpAlaLeuPheAspSerThrLysGluMetAlaTrpIleProLeu 427
Db 1297 TCCGCGCTTCGATGTGAGAGCTG-----ATGCTGCGGTGTTCGTC 1338
Qy 428 AspGlnGlnGlnLysValLysGlyLysAlaAlaIleAspValAlaAlaArgPheGly 447
Db 1339 CCAGATTTCCCTGCTGTGCTGATTAACATCAATTCATCAGGACGATTCGCGGTGGT 1398
Qy 448 LysSerGlyLysAlaLeuIle-----GlnGlnGlyLeuLeuValIleCysGlySer 464
Db 1399 GACTCGGCAACTCAGGATGATCACACCGCTGATCCGTAATGATGTGATCTTCGCTTGG 1458
Qy 465 IleGlyAlaMetThrPro-----TyrLeuAlaValIleLeuLeuPhe 478
Db 1459 CTCCTCGATACAGACCGATGATGATTTGAGCACTTGAATGTCACAGCTTATCCATT 1518
Qy 479 IleIleAlaIleTrpLeuValSerAlaThrLysLeuAsnLysLeuPheAlaGlnSer 498
Db 1519 GTATCCCTTTCGTGGCTAGCGCTGAGCTACATTTGTGCAATTTGTCTACCGCGATTTG 1578

```



```

QY      408 SerLysSerThrLysTyrAlaLeuPheAapSerThrLysGlnMetAlaTyrLLeProLeu 427
      113766 TCCGGCTCTCGCATGTGACGCTG-----ATGGCTGGGGTGTTCGT 113807
QY      428 AepGlnGluGlnLysValLysGlyLysAlaAlaIleAspValAlaIleAaTgPheGly 447
      113808 CCGATGTTGCTCTGTGCTTGGCTATGAACATATCTTATTCAGGACCATTCGGCGTGGGT 113867
QY      448 LysSerGlyGlyAlaLeuIle-----GlnGlnGlyLeuLeuValIleCysGlySer 464
      113868 GACTGGGCACTCAGGTGATACACCCGCGTAATCCGTACATGTTGTGATGCTTCGGTTTG 113927
QY      465 IleGlyAlaMetThrPro-----TyrLeuAlaValIleLeuLeuPhe 478
      113928 CTCCTCGCATAGCAACCGGATGCAGGTTTAAAGCACTTGATGTGACAGCTTATTCACATT 113987
QY      479 IleIleAlaIleTyrLeuValSerAlaThrTyrLysLeuAsnLysLeuPheLeuAlaGlnSer 498
      113988 GTGATCCCTTTCCTGGCTTACCGCTGATCATATTGTTGGCAATTGTTGTTCAAGCGCATTTG 114047
QY      499 AlaLeu 500
      114048 CCGCTT 114053
Db
RESULT 8
US-09-815-242-6064
Sequence 6064, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlgen, Kari L.
APPLICANT: Zykend, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011a
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6064
LENGTH: 1389
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1389)
US-09-815-242-6064

Alignment Scores:
Pred. No.: 1,34e-05 Length: 1389
Score: 138.50 Matches: 111
Percent Similarity: 37.55 Conservatvity: 70
Best Local Similarity: 23.03% Mismatches: 172

```

| | | | |
|--------------|-------|---------|-----|
| Query Match: | 5.27% | Indels: | 130 |
| DB: | 9 | Gaps: | 24 |

US-09-869-433-2 (1-515) x US-09-815-242-6064 (1-1389)

| | | | |
|----|-----|---|------|
| QY | 69 | TTpLeuValValProCysAlaIle-----IlePheMetLeuIleTyrAlaLys | 84 |
| | | | |
| DB | 109 | TGGATATACAGTACGATGTCCTACTCTTGCTCTCTGTCGTGATGCTAATTTAGCGAGATT | 168 |
| QY | 85 | LeuSerAenIle-LeuSerLysGlnAlaLeuPheTyrAlaValGlyThr----- | 1000 |
| | | | |
| DB | 169 | ACCGTTAATCTCAATAAATACGGTTTATTTATTTACTACCGATCAACTCTTTTATTAAAC | 228 |
| QY | 101 | -----ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuAr | 117 |
| | | | |
| DB | 229 | GCATTACCCCTCGCTTTCGGCGCATTTATGTCGTCTCCACTCTCTTTATGCGCTATA | 288 |
| QY | 117 | GAAPrAlaLeuHisProThrGluPheAlaAspPrArgLeuGlnAlaIleLeuProProGlyLe | 137 |
| | | | |
| DB | 289 | TTTCGTGGACCGCCGATGGACGGTTTTT-AGTACTGCAATCTGATTTATTTCTTGGCTCG | 347 |
| QY | 137 | uLeuGlyLeuValAlaIleLeuAArgAsnTrpThrPheAlaAlaPheTyrValLeuAlaG | 157 |
| | | | |
| DB | 348 | GCTCGGAATTCGCCGCAAAATCCGAATACCTCTTTGGGATATTATATCGTTATCGCTTT | 407 |
| QY | 157 | uLeuTrpGlySerValMetLeuSerLeuMetPheTrpGly-----PheAlaAsnGluI | 175 |
| | | | |
| DB | 408 | GCTATGCGCGT-----TTTCGACGGTGCAAACTTGTCTTCGAGCAT | 446 |
| QY | 175 | eThrIysIleHis-----GluAlaIysArgPheTyrAlaLeuPheGlyIleG | 191 |
| | | | |
| DB | 447 | GCGCAATATACGTTTCTCTTCCAAAAGCCAAAGCGAGCGCTCTT--GGGATTTAA | 503 |
| QY | 191 | YAlaAsnIleSerLeuLeuAlaSerGlyArgAlaIleValTrpAlaSerLysLeuArgAl | 211 |
| | | | |
| DB | 504 | TGGCGGATTA-----GGAACCTTAGCTG | 527 |
| QY | 211 | AserValSerGluGlyValAspProTrpGlyIleSerLeuAArgLeuMetAlaMetTh | 231 |
| | | | |
| DB | 528 | AAGTGTATATGACAGCTGTTGACCGCTGTCATTTTGGACGTATTTGGCTTTCTCGG | 587 |
| QY | 231 | rIle-----ValSerGlyLeuValLeuMetAlaSerTyr | 242 |
| | | | |
| DB | 588 | CGTCATATGGCGTACCGCAGCGCAGCGATGCTGATGTCGTGGCCGATATGCCGATGAT | 647 |
| QY | 242 | rTrp-----TrpIleLeuLysAsnValLeuTh | 251 |
| | | | |
| DB | 648 | TTGGGTACCGCTACTGCGGATGGCCACAGATCCGCCGATGTCAGAGGATGATATATATC-- | 705 |
| QY | 251 | rAspProArgPheTyrAsnProGluGluMetGlnLysGlyLysGlyAlaLysProLys | 271 |
| | | | |
| DB | 706 | -----GCCAGTTCACG | 716 |
| QY | 271 | sMetAsnMetLysAspSerPheLeuTyrLeuAspArgSerProTyrIleLeuLeuLeuTh | 291 |
| | | | |
| DB | 717 | CGCCTCATATGGCCGACAGCTCCCTGTCTTACACAGCGCTG--CANCCTCGGCTCGAG | 773 |
| QY | 291 | rLeuLeuValIleAla--TyrGlyIleCysAlaIleAsnLeuIleGluValThrTrpLysSe | 310 |
| | | | |
| DB | 774 | CTGTGTTTACCTTGCCACCTTCGTGCTGTTATTCGGT--TTTTCTGCGGGTTTTCGAT | 830 |
| QY | 310 | rGlnLeuLysLeuGlnTyrProAsnMetAsn-----AspTyrSerGluP | 325 |
| | | | |
| DB | 831 | GCTGGCAAAAACCCGATTCGCCGATGTGAATTTCTGCGCCTTCGCTTTGGCCCAT | 890 |
| QY | 325 | eMetGlyAsnPheSerPheTrpThrGlyValValSerValLeuIleMetLeuPheValG | 345 |
| | | | |
| DB | 891 | TATCGGT-----GCCATTCGCGCGTTCGGTGG | 917 |
| QY | 345 | YGLYAsnValIleArgLysPheGlyTrpLeuThrGlyAlaLeuValThrProValMetVa | 365 |
| | | | |
| DB | 918 | TGGTGCTATTTCCGATTAAGTTCGGCGCGGTGCGGGTGACGTTGATTCACCTTATTTTAT | 977 |

```

Score: 138.00 Matches: 25
Percent Similarity: 76.74% Conservative: 8
Best Local Similarity: 58.14% Mismatches: 10
Query Match: 5.25% Indels: 0
DB: 11 Gaps: 0

US-09-869-433-2 (1-515) x US-09-770-961-833 (1-484)

QY 16 LEUTPRPROILHISTHRISGLILEULYLYLVALLEUPROMETPHELEMETPHEPHE 35
      :::::|||||
Db 335 ATTTCGGGTGGAGGTTACGACTCTGAAGAAATGTTCTTTAGGGCTNATGTTCTTT 414
      :::::|||||
QY 36 CYSILERTHPHEASNLYRTHRVALLEUARGLASPRTLYLSAPTHRLLEULVALGLYALA 55
      :::::|||||
Db 415 TGCAATCCITTTCAATTACCAATCTTTAGGGAACAGAGATGTTTGGTGTACGGCT 474
      :::::|||||
QY 56 PROGLYSER 58
      |||||
Db 475 AAAGGAGCT 483
      |||||

RESULT 10
US-10-329-960-1/c
; Sequence 1, Application US/10329960
; Publication No. US2003009277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fragi
; TITLE OF INVENTION: Theoreof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02

```

```
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10091)..(10091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
```

```

FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152530)..(152530)

Alignment Scores:
Pred. No.: 4.62 Length: 1830121
Score: 133.00 Matches: 120
Percent Similarity: 35.42% Conservative: 89
Best Local Similarity: 20.34% Mismatches: 158
Query Match: 5.06% Indels: 224
Gaps: 30
DB: 14

US-09-869-433-2 (1-515) x US-10-329-960-1 (1-1830121)

QY 15 PheLeuTrpProIleHisThrHisGluLeu-----LysValValLeuPro 29
DB 1779409 TTTATTGGCCAGCAAGCAATCAAGCATTATGATCAAGCTAAAGTGATATTTTGCT 1779350
QY 30 MetPheLeuMetPhePheCysIleThrPheAsnTyThrValLeuArgAspThrLysAsp 49
DB 1779349 AATTTTACTGGTTTATGACTTACTTCTTTCAGTG----- 1779314
QY 50 ThrLeuIleValIglValAProGlySerGlyValAglValIleProPheIleLysPheTrp 69
DB 1779313 -----TTTTAGCATTTTGA 1779299
QY 70 LeuValValProCysAlaIleIlePheMetLeuIleTyAlaLysLeuSerAsnIle--- 88
DB 1779298 TTAATTT-----TTATCATGATGATGACTTGGCAATATATAA 1779263
QY 89 -----LeuSerLysGlnAlaLeuPheTyR 96
DB 1779262 CTAGGCGCAAGATGAAGAAACCTGAATTTAGTTCTTATCTTGGCTTGCATGATATTT 1779203
QY 97 AlaValGlyThrProPhe---LeuIlePhePheAlaLeuPheProThrValIleTyPro 115
DB 1779202 GCCCGTGAATGGGGGTTGGGCTGATGTTTTTGGCGTAGCAAGAACATTAACCATAT 1779143
QY 116 LeuArgAspValIleuHisProThrGluPheAlaAspValGluGlnAlaIleLeuProPro 135
DB 1779142 CTTTCTGACATTACAGCAGGTTCTGCAGAACATTAACAAACAAGAGCTTTACTT----- 1779089
QY 136 GlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThrPheAlaIlePheTyValLeu 155
DB 1779088 -----CATACTCTGTTCCACTGCGGAAATTCACGCGTGCGGACGTA--- 1779030
QY 156 AlaGluLeuTrpGlySerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIle 175
DB 1779049 -----TATGGCACCATTTGCTTAGCATTAAGCTTATTTGGGTTTCGTTATATAATTA 1778999
QY 176 Thr-----LysIle 178
DB 1778998 CTTTAGCGTTCGCTCTTGTGTTTTTATCCTTATTAAGATCGTATTAATGCAAAATC 1778939
QY 179 HisGlnAlaLysAspPheTyAla-----LeuPheGlyIle----- 190
DB 1778938 GCGCATGCAATTAATGATTATGCGCTTACTTGCACATTAATTTGGATTAATTAACACATTA 1778879
QY 191 GlyAlaAsnIleSerLeuLeuAlaSerGly---ArgAlaIleValIleTrpAlaSerLysLeu 209
DB 1778878 GGTTTTGGTTCATCAAGCTTGGAGGCTTGAAACAAATAGGTGATTAAGTCAGAAC 1778819
QY 210 ArgAlaSerValSerGluGlyValAspProTrpGlyIleSerLeuArgLeuLeuMetAla 229
DB 1778818 AGCTTTGGCTTACAAAGTTGGGTTATTT-----GTTGCGTAGATGTTTATGCGGTG 1778768
QY 230 MetThrIleValSerGlyLeuValLeuMetAlaSerTyTrpTrpIleAsnLysAsnVal 249
DB 1778767 TTTTCTGCTATTTCTGGTGT----- 1778747
QY 250 LeuThrAspProArgPheTyAsnProGluGluMetGlnLysGlyLysGlyAlaLys 269
DB 1778746 -----GGAAAGGCGGTGAAA 1778732

```

```

QY 270 -----ProLysMetAsnMetLysAspSer-PheLeuTy----- 280
DB 1778731 ATATTAAAGTAATAATTAACCTTAAGCATTATGTTACTCTCTTGTGTTTAATTCA 1778672
QY 281 -----LeuAspArgSerProTyIleLeuLeuLeuThrLeuValIleAlaTyGlyIle 299
DB 1778671 GGGCTTCGTTATACCTTTATATCGCATTTAGGACAAATATGTTATTT-----AT 1778621
QY 299 eCysIleAsnLeuIleGluValThrTyPlySerGlnLeuLysLeuGlnTyRProAsnMe 319
DB 1778620 TTCAGC-AAATTAGTCGACACTGATTAAAC-----TATGCTTATGA 1778577
QY 319 tAsnAspTySerGluPheMetGlyAsnPheSerPhe----- 331
DB 1778576 ACAAGAAACATACACTGCTGTTTACCGGATGAGCTGCTTATTTGGCGTGTGCTTC 1778517
QY 332 -TrpThrGlyValAlaSerValLeuIleMetLeuPheValGlyValAsnValIleArgLys 351
DB 1778516 TTGGGCTCCGTTTGTGGGTTTATTTATTTAGCGCGTATTTCTAAAGGGGGAACATACCGTGA 1778457
QY 351 sPheGlyTyRLeuThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleVal 371
DB 1778456 ATTT-----ATTTTGGCGTATTAATTAATCA-----AGTTTATTTGGTATTTT 1778412
QY 371 lPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyTh 391
DB 1778411 ATGGTT-----ACTGTTTGGTAA 1778391
QY 391 rThrProLeuMetLeu-----AlaValValAlaGlyAlaIleGlnAsnIleLeuSerLys 409
DB 1778390 TACAGCAGTATGGCTAATATGATGSCATTCGCGGGGCGGCGTGGCAATTAATTTCT- 1778333
QY 409 sSerThrLysTyRAlaLeuPheAspSerThrLysGluMetAlaTyRTrleProLeu----- 427
DB 1778332 -TCCCGCAAAATTTTATTAATTT-----AAATTTTAATTAATTCGCTTAAACCAAC 1778283
QY 427 ----- 427
DB 1778282 AATAACAGCGCTTGTGACTTATTAATTAATTTATTTGTTTATATCACTTACGCGGATTC 1778223
QY 428 -----AspGlnGlu----- 430
DB 1778222 AGATATTATGTTGTTAATAATCAATTCGATCTCGGATTAAGTTAGCTTGCCTGCGGTG 1778163
QY 431 -GlnLysValLysGlyLysAlaAlaIleAspValAlaAlaArgPheGlyLysSerGly 450
DB 1778162 GCAAGCATATATGGGGAACCTTAATGTCGCTGTTGCGATTTGCTTATGCAATCTGG 1778103
QY 450 YGly---AlaLeuIleGlnGlnGlyLeuLeuValIleCysGlySerIleGlyAlaMetTh 469
DB 1778102 TGGACTTGCTAATTTGCAACATGACATTAATTTGTC-----TT 1778061
QY 469 rProTyRLeuAlaValIleLeuLeuPheIleIleAlaIleTyR-----LeuValSer-- 486
DB 1778060 GCCTTTGCCTTAATGATGTTGGAATGCTTTTATGATTAAGAAAGCTTAATAGCGGA 1778001
QY 487 -----AlaThrLysLeuAsn 491
DB 1778000 TAAAAAATATTTTCTACTATAAGTCAT 1777973

RESULT 11
US-09-815-242-6047
; Sequence 6047, Application US/09815242
; Patent No. US2002006159A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

```



```

; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Ircvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-014-717-1

Alignment Scores:
Pred. No.: 0.437 Length: 68750
Score: 122.50 Matches: 105
Percent Similarity: 34.68% Conservative: 67
Best Local Similarity: 21.17% Mismatches: 165
Query Match: 4.66% Indels: 159
DB: 13 Gaps: 21

US-09-869-433-2 (1-515) x US-10-014-717-1 (1-68750)
QY 76 IletlePheMetIleuTyraIalysleuSerAnilleuSerIySglN----- 92
DB 3460 CTCGGGCTCATCTCTGTACCGCGCGCTCCGCGAGCTCGCGCGCGCGCCAG 3519
QY 93 -----AlaleuPheTyraIalysIlyThrProPheIleuIlePhePheAla 107
DB 3520 CCCGAGGTCTCGGGAGGCTTTCCGCGCGCTC-----GTG 3555
QY 108 LeuPheProThraIleTyraIleuThraPheValIleuHisProThraIleuAlaasp 127
DB 3556 CTGGGCCCCCTCGCTCGCGCGCGCTCGCGCGCGGTTTCATCGAGCCCTCTCCAGAG 3615
QY 128 ArgIleuGlnAlaIleuProGlyIleuLeuGlyIleuValAlaIleuArgAsnTrp 147
DB 3616 CCGCGGCTCGGGGTCTGCTCGGCGCATC----- 3645
QY 148 ThrPheAlaAlaPheTyraIleuValIleuTrpGlySerValMetIleuSerIleuMet 167
DB 3646 -----TCTCGATAGCGCGCTCTCTCTGCTG 3675
QY 168 PheTrpGlyPheAlaAsnGluIleThraIleHisGluAlaIlyArgPheTyraIleu 187
DB 3676 ATGGCGGCGCATCGAGTGTGGCATCTCGCGCAAGAGCGCGCGCGCGCGCTC 3735
QY 188 PheGlyIleGlyAlaAsnIleSerIleuAlaSerGlyArgAlaIleValTrpAlaSer 207
DB 3736 TCGGCGCTCGCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3780
QY 208 LysIleuArgAlaSerValSerIleuGlyValAspProTrpGlyIleSerIleuArgIleu 227
DB 3781 ---TTCGCGCGCTGCTCTGATCGAGCGCGCTTCGAGCGCGCTTCTCGGAGTCTG 3837
QY 228 MetAlaMetThrIleValSerIleuVal-----LeuMet 239
DB 3838 CTCCTCGGAGCGCGCGCATCGGTATCCGAGGTCTGATCGAGCGCGAGTCTGATGCGC 3897
QY 240 AlaSerTyraIleuTrpIleAsn-----LysAsnValIleuThraAspProArgPheTyraAsn 257
DB 3898 CGCAGCTTATGCGCGAGGTACGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3957
QY 258 ProGluMetGlnIlySglIyAlaIyProIyMetAsnMetIyAsnSer 277
DB 3958 CTCGTCGCGATGACGTCGTCGAGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4017

```

```

QY 278 PheIleuTyraIleuAspArgSerProTyraIleuIleuThraIleuValIleAlaTyra 297
DB 4018 CGGCTCTCG-----GGAGCGGATTTCTGCTGTTCATGAGTCTCTGTC----- 4059
QY 298 GlyIleCysIleAsnIleuIleGluValThrTrp-----Lys 309
DB 4060 GGGGCGGCGCTCACCGACCTCGCATGGCTGGGTGGCGCGAGCGCGCGCTCTCCAG 4119
QY 310 SerGlnIleuTyraIleuGlnTyraIleuProAsnMetAsnAspTyraIleuPheMetGlyAsnPhe 329
DB 4120 GCAGACGATGCTGCTC----- 4134
QY 330 SerPheTrpThrGlyValIleValSerValIleuIleMetIleuPheValGlyIyAsnValIle 349
DB 4135 -----GTCCTCTCTCTACAGTTCTCTGCGCGCGCGCGCTGACG 4170
QY 350 ArgIySphGlyTrp-----LeuThrGlyAlaIleuValThrProValMetVal----- 365
DB 4171 CAGCGGCTCGGCTCGACCGCGCTGCTGGCGCGTTCGCGCGCTGCTGCTCAAGC 4230
QY 366 -----LeuIleuThrGlyIle----- 370
DB 4231 GCTCTCGCACCAACCGCTCTCTCTGACGCGCTGCGAGACGCTGCGCGCGCTCTTC 4290
QY 371 -----ValPhePheAlaIleuValIlePheArgAsnGlnAlaSerGlyIleu-----Val 386
DB 4291 CGCGCTGTGTTCTTCTGCTCTGCGCGCATGCGCTCGACGTTCTGCACTGCGACCGCG 4350
QY 387 AlaMetPheGlyThrThrProIleuMetIleuAla-----ValValVal 400
DB 4351 CGCGCGTGGGAGCGTGTGCTGCTGCGCGACCGCGAGCGCGGAGGTGTCGCC 4410
QY 401 GlyAlaIleGlnAlaIleuSerIySerThraIyTyraIleuPheAspSerThraIy 420
DB 4411 GCCCGCTCGCGCGCGCGCTCGCG-----GGCTCAAGGCGAGCGAGCGCG 4455
QY 421 GluMetAlaTyraIleProIleuAspGlnIleuValIySglIyAlaIleAlaIleAsp 440
DB 4456 CGGCTCGTGGCGGTGCGCTGACATG-----AAGCGCGGACGAGC 4497
QY 441 ValValAlaAlaArgPheGlyIySerGlyAlaIleuIleGlnIleuIleuVal 460
DB 4498 CTACATCGCGCATCGCGCG-----GTGCGAGCTCGGCTCTCTC--- 4536
QY 461 IleCysGlySerIleGlyAlaMetThrProTyraIleuAlaValIleuIleuPheIle 480
DB 4537 -----TCCACGAGGCTTATACATGACCGCGCTGCGCGTGCACGCTGAC 4587
QY 481 Ala-----IleTrpIleuValSerAlaThrIySleuAsnIySleuPheIleu 495
DB 4588 GCTTCACCGCGCGCTCTCATCTGCTGAC----- 4617
QY 496 AlaGlnSerAlaIleuTyraIleuGlnIleuValAlaGlnIleuAspSerAla 511
DB 4618 -----AAAGGCGCGCTCGACGAGGAGGTGCGCT 4650

RESULT 13
US-09-815-242-9794
; Sequence 9794, Application US/09815242
; Patent NO. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zykkind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A

```



```

CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9794
LENGTH: 1188
TYPE: DNA
ORGANISM: Salmonella typhi
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1188)
US-09-815-242-9794

```

```

Alignment Scores:
Pred. No.: 0.00127      Length: 1188
Score: 120.50          Matches: 98
Percent Similarity: 34.55%      Conservative: 72
Best Local Similarity: 19.92%    Mismatches: 189
Query Match: 4.58%             Indels: 133
DB: 9                        Gaps: 20

```

US-09-869-433-2 (1-515) x US-09-815-242-9794 (1-1188)

```

QY 28 LeuProMetPheLeuMetPhePheCysAlIlePhePheAsnYrThrValLeuArgAspThr 47
DB 7 TTATGATGTCCTCGCTGCTTATTTGCACTGACATAAGTGGCTTCGCTATT--GGCAGC 63
QY 48 LysAspThrLeuIleValGlyAlaProGlySerGlyAlaGluAlaIleProPheIleLys 67
DB 64 ACTGAGTTCGTAATGTCGCGC-----CTGGTCCGACCATTCGCA 102
QY 68 PheTrpLeuValVal-----ProCysAlaIleIlePheMetLeuIleTyrAlaLysLeu 85
DB 103 CAACAACCTGCTGCTGCTTACCTTCTGCGGAATGCTGTTTATCTAT----- 153
QY 86 SerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThrProPheLeu----- 103
DB 154 -----GCACTGGGTGGCTATTGGCGCGCGCTGCTGACCGCT 192
QY 104 -----IlePhePheAlaLeuPheProThrValIleTyr 114
DB 193 CTGACCGGAGCGCTGCGCGCTGAACAGTTACTGCTGCGCTGAAG-----GTTTGT 246
QY 115 ProLeuArgAspValLeuHisProThrGluPheAlaAspArgLeuGlnAlaIleLeuPro 134
DB 247 ACG-----GCGGGGAATTGCTGCGCGTGGCAGGCT 276
QY 135 ProGlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThrPheAlaIlePheTyrVal 154
DB 277 CCGGGCTATATGACATTTATATGCGC--GCCGCTGCTGACGAGGCGCTGGCGCAGCGGCTG 333
QY 155 LeuAlaGluLeuTrpGlySerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGlu 174
DB 334 TTCTTCGATTTGGCTCTACCATCGCAAGCTG----- 369
QY 175 IleThrLysIleHisGlyAlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIle 194
DB 370 GTGCCGAAAGAGAAAGCGCTTCTGCTATCGCTATTTATGTTGCT--GGTTTAAAGTA 426

```

```

QY 195 SerLeuLeuAlaSer-----GlyArgAlaIleValTrpAlaSer 207
DB 427 GCGCTGTACAGAGTGTCCGCTGGGACATTTATCGTACAGATTGGCTGCGC--- 483
QY 208 LysLeuArgAlaSerValSerGlyValAlaAspProTrpGlyIleSerLeuArgLeuLeu 227
DB 484 -----GAGACCTTC 492
QY 228 MetAlaMetThrIleValSerGlyLeuValLeuMetAlaSerYrTrpTrpIleAsnLys 247
DB 493 CTGGCGTATCACCTGATGAGCGGTGATGCGCTAGTAGAGTCACTGATGCTGCTGCC 552
QY 248 AsnValLeuThrAspProArgPheTyrAsnProGluGluMetGlnLysGlyLysGly 267
DB 553 AATATT-----CCTGTCGGCGGCC 573
QY 268 AlaLysProLysMetAsnMetLysAspSerPheLeuTyrLeuAspArgSerProTyrIle 287
DB 574 GCC-----ACATTCGTGACAGGTGAAAGTACTGACTCATCTCGTCTGCTG 621
QY 288 LeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCysIleAsnLeuIleGluValThr 307
DB 622 CTGATTTACCGCCGTTACGGCGCTGCGCTACGCGGGGTATTC----- 663
QY 308 TrpLysSerGlnLeuLysLeuGlnTyrProAsnMetAsnAspTyrSerGluPheMetGly 327
DB 664 -----ACGCGATTACCTTCCTGCGCGCGCATGATGACAGATTCGCGGATTTCCTCG 717
QY 328 AsnPheSerPheTrpThrGlyVal-----ValSerValLeuIleMetLeuPheVal 344
DB 718 GCAAGACATAAGCTGATCTCCTACCTGAGGTATGAGGTATCGAGTATCGCAATATCTGG 777
QY 345 GlyGlyAsnValIleArgLysPheGlyTyrPheLeuThrGlyAlaLeuValThrProValMet 364
DB 778 GCGCGTAACCTGGCGGATGAAGCATGCT-----GCCGTCCGCGCTCTG 819
QY 365 ValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGly 384
DB 820 AATTCATTTCGCGCGCGCTGTTTCTCTGCTGATGATGCTTCCAGAGTACGCTCCACG 879
QY 385 LeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValAlaGlyAlaIleGln 404
DB 880 CAGTATCCCGCGCTGCTACCATTCGTGATGGAGTCTTTCGCTCGGAACGTACCG 939
QY 405 AsnIleLeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLysGluMetAlaTyr 424
DB 940 GGGTTG-----CAGGCTTAC 954
QY 425 IleProLeuAspGlnGluGln-----LysValLysGlyLysAlaAlaIleAsp 440
DB 955 GTGGTACAGAAAGCGGAACAATTACAGCCTAATGCGTTAATGTGGCGTCAAGCTTAAC 1014
QY 441 ValValAlaIleAlaArgPheGlyLysSerGlyValAlaLeuIleGlnGlnGlyLeuVal 460
DB 1015 ATTCCGCGCTTATACATGATGATTTGCGCTGCGTTCCTGATTTGGCGGACAAAGCTGGCG 1074
QY 461 IleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValIleLeuLeuPheIleIle 480
DB 1075 CATTAACGTTTG-----GCGCAGACGCGGTGATGGCCCTCATGTTCTGTGGTGGCT 1128
QY 481 AlaIleTrpLeuValSerAlaThrLysLeuAsnLys 492
DB 1129 TTCTCTGATGAGCGCTGACGAGACGCTTGATTA 1164

```

RESULT 14

```

US-09-815-242-8073
Sequence 8073, Application US/09815242
Patent No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haeselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel

```

APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO: 8073
 LENGTH: 1278
 TYPE: DNA
 ORGANISM: Staphylococcus aureus
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)...(1278)
 US-09-815-242-8073

Alignment Scores:
 Pred. No.: 0.00142 Length: 1278
 Score: 120.50 Matches: 94
 Percent Similarity: 37.27% Conservative: 86
 Best Local Similarity: 19.46% Mismatches: 160
 Query Match: 4.58% Indels: 143
 DB: 9 Gaps: 19

US-09-869-433-2 (1-515) x US-09-815-242-8073 (1-1278)
 QY 9 PheGlyLysLeuArgSerPheLeuThrProIleHisGluLeuLysValLeu 28
 DB 19 TTGGAGCGCGAAATTAAATATCCGCCCATGCTTGCCATACACGCGGTCAAAATATG 78
 QY 29 PrometPheLeuMetPhePheCysIleThrPheAsnTyrThrValLeuArgAspThrLys 48
 DB 79 TGGATGTGATGCTAGGCTTTCCTTACA-----GGCATATTACTCCCTTTATT 129
 QY 49 AspThrLeuIleValGlyAlaProGlySerGlyAlaGluAlaIle----- 63
 DB 130 ACTGTTATGTTGTTGATTTTATGATGAGGTGTAAGGTGTAAGGTGTAAGGTGTAAGGT 189
 QY 64 ProPheIleLysPheThrLeuValAlaProCysAlaIleIlePheMetLeuIleTyrAla 83
 DB 190 CCATGCTTCGGGTTATTTTGGCTGTC-----GTGATTACATGCTTATCGGACGA 240
 QY 84 -----LysLeuSerAsnIleLeuSerLysGluAlaLeuPheTyrAlaVal 98
 DB 241 TTTTACGATATTCACGCTGCTGCAATATGTCGG-----TACGAATAT 282
 QY 99 GlyThr-----ProPheLeuIlePhePheAlaLeu 108
 DB 283 GGTACAAACACATTTTACTGTCGACACCAATGATGATTAATATTCGACGACGTC 342
 QY 109 PheProThrValIleTyrProLeuArgAspValLeuHisProThrGluPheAlaAspArg 128
 DB 343 TTTTGGCATGCTTACTGATTAAGT-----TTAAATCATGCAAAATGCTTGATTAAT 396
 QY 129 LeuGluAlaIleLeuProGlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThr 148

DB 397 TTAGTAATATTAATACCGTATTTACTATTAATGTCGCTATTA-----AGT 447
 QY 149 PheAlaAlaPheTyr----- 153
 DB 448 ATTGCTGTCAATTTCAACCCGTGAATCTGACCTAGTACCTAGATTAATATATACA 507
 QY 154 -----ValLeuAlaGluLeu 158
 DB 508 CATCCTTCATTTGAGGAAGTTGGAAGGCTATTTACAAATGACCTGTGTGCGCTTA 567
 QY 159 TrpGlySerValMetLeu----- 164
 DB 568 GCTTTTCCGAGTACATGTCATGCTATTAAGTTAAAGGCTTCACAGATCCGACGAAA 627
 QY 165 SerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleHisGluAlaLysArgPhe 184
 DB 628 ATATTAAAGTATGCTCTCTTTCAGATTATTCACCTATTAATCTGATGATTATAC 687
 QY 185 TyrAlaLeuPheGlyIleGlyAlaAsnIleSer-----LeuLeuAlaSerGlyArgAla 202
 DB 688 TTTCGTTGGCATATGATGATGCTGTCACAGCTCAGAAACTTAAGATGATCTGAT 747
 QY 203 IleValTrpAlaSerLysLeuArgAlaSerValSerGluGlyValaAspProTrpGlyIle 222
 DB 748 ATATTGACATACATTCATTAACGGTATTTGGTTGCTCGGTAACTCGATTTGGAATG 807
 QY 223 SerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAlaSerTyr 242
 DB 808 ACGGTT--ATTCTTGATGCTCAACACATGATATGACCTGCTC--AAAGCTTGTCGA 861
 QY 243 TrpTrpIleAsnLysAsnValIleThrAspProArgPheTyrAsnProGluGluMetGln 262
 DB 862 ACATTACTAAGAACACGTC----- 882
 QY 263 LysGlyLysLysGlyAlaLysProLysMetAsnMetLys----- 275
 DB 883 -----CTTAAGTTTCTTAATAAAATATTCGACCTGTTTCTCT 921
 QY 276 ---AspSerPheLeuTyrLeuAspArgSerProTyrIleLeuLeu-----LeuThr 291
 DB 922 ATCATAGATTTTATTTTACACACTTGTTAGAAATGATTTTAAATATGCTGTAACA 981
 QY 292 LeuLeuValIleAlaTyrGlyIleCysIleAsnLeuIleGluValThrTrpLysSerGln 311
 DB 982 TTGTTGACATTAATATATCCCGTCTATTTGCACTTACTGATATATATTT----- 1032
 QY 312 LeuLysLeuGlnTyrProAsnMetAsnAspTyrSerGluPheMetGlyAsnPheSerPhe 331
 DB 1033 -----GCTAACATGTTTAGACATTCAGATTCAAGTTGGGCTAT 1071
 QY 332 TrpThrGlyValAlaSerValLeuIleMetLeuPheValGlyLysAsnValIleArgLys 351
 DB 1072 CGACTCGCAACTGTTATTTACATGATTAATTTCAATTTTA-----CAAAATCAAAATAGT 1125
 QY 352 PheGlyTrpLeuThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleVal 371
 DB 1126 TTCACTTAATTAACGGGTGTTATTTGAAATGTTTATGATGTTTACTCTTAAGCAGATATC 1185
 QY 372 PhePheAlaLeuValIlePheArgAsnGluAlaSerGlyLeuValAlaMetPheGlyThr 391
 DB 1186 GATTTACTTGGCTGTA----- 1203
 QY 392 ThrProLeuMetLeuAlaValAlaValGlyAlaIleGluAsnIleLeuSerLysSerThr 411
 DB 1204 ---CAATTCAATCTTTTGGCAATCATGTTTCAATCATCATGATGATTAATACCGCTCG 1260
 QY 412 LysTyrAla 414
 DB 1261 AAACAGCG 1269
 RESULT 15
 US-10-085-959-109/c

```

Sequence 109, Application US/10085959
Publication No. US20030165870A1
GENERAL INFORMATION:
APPLICANT: Blactner, Frederick R.
APPLICANT: Welch, Rodney A.
APPLICANT: Butland, Valerie D.
TITLE OF INVENTION: No. US20030165870A1 Sequence of E. Coli CFT073
FILE REFERENCE: 960296.97648
CURRENT APPLICATION NUMBER: US/10/085,959
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: 60/242,412
PRIOR FILING DATE: 2000-10-19
NUMBER OF SEQ ID NOS: 255
SOFTWARE: PatentIn version 3.1
SEQ ID NO 109
LENGTH: 76804
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: misc feature
LOCATION: (39751)..(39751)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (49372)..(49372)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (66539)..(66539)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73693)..(73693)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73705)..(73705)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73737)..(73737)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73759)..(73759)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73761)..(73761)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73766)..(73766)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73774)..(73774)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73781)..(73781)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73798)..(73798)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73802)..(73802)
OTHER INFORMATION: Unsure
US-10-085-959-109

Alignment Scores:
Pred. No.: 0.886 Length: 76804
Score: 120.50 Matches: 109
Percent Similarity: 36.65% Conservative: 79
Best Local Similarity: 21.25% Mismatches: 173
Query Match: 4.58% Indels: 152
DB: 12 Gaps: 27

US-09-869-433-2 (1-515) x US-10-085-959-109 (1-76804)
CY 94 LeuAeYrAAlVaAlGyTnPrPhLeuIleHehehAlaLeuPhProthrValIle 113
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
71399 AATATTCGATCGGCGGACGACGATATTACTTGACGATTTGATTGTCGATATCT 71240

```

| | | | |
|----|-------|---|-------|
| QY | 114 | TyrProLeuMkrgsPValLeuNHsPROMThguPhenalaAspArgLeuGlnAlaIleLeu | 133 |
| Db | 71239 | CTGGTCTTTCACAGAGGTTCAGCAT-----GAATTCGGC-----TTGACGACAAATAGAA | 71192 |
| QY | 134 | ProProGlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThrPheAlaIaPheTyr | 153 |
| Db | 71191 | GCAGCAAGCCATTATTCGCCGTGCTTTTATTTCCGATGGTTGGGGGGCTGTGATTGGC | 71132 |
| QY | 154 | ValLeuAlaGluLeuTrpGly-----SerValMetLeuSerLeuMetPheTrpGly | 170 |
| Db | 71131 | GCATTATTCGACAAAGATGGGGCCACAGATGGCAATGGATTGTTTAAGTTATAGTACG----- | 71083 |
| QY | 171 | PheAlaAsnGluIleThrLysIleHisGluAlaIaArgPheTyrAlaLeuPheGlyIle | 190 |
| Db | 71080 | -----CTATTCTCTCTC----- | 71065 |
| QY | 191 | GlyAlaAsnIleSerLeuLeuAlaIaSerGlyAArgAlaIleValTrp----- | 205 |
| Db | 71068 | GGAACCCCTTCCTGCTGGCTGGCTCCTCGGATATGACAGTATGTTTATTCCTCGCATTTGT | 71009 |
| QY | 206 | -----AlaSerLysLeuArgAlaIaSerValSerGluIaIaValaAspProTrp | 220 |
| Db | 71008 | ATTGGGCTTGGTATGGCAGGTGATATGGCTCAAGTGTACATATTTATTCGAAGCTGG | 70948 |
| QY | 221 | GlyIleSerLeuArg-----LeuLeuMetAlaMetThrIleValSer | 234 |
| Db | 70948 | CCTGTTCATCTCAGAAACAGGCAAGTGGTTTATTTATTTCCGAAATTTCTATTGGGGGA | 70888 |
| QY | 235 | GlyLeuVal-----LeuMetAlaSerTyrTrp----- | 243 |
| Db | 70888 | GGTTTACTCTCAGACGTTATTATGACATTGTATGACACTATGGGGCTGGCTCTCTTTTC | 70829 |
| QY | 244 | -----TrpIleAsnLysAsnValLeuThrAsp | 252 |
| Db | 70828 | TTTTCGGGCATGCTCCAAATATTATTGCTTTTATCTTCGCAAGATCTT----- | 70778 |
| QY | 253 | ProArgPheTyrAsnProGlu-----GluMetGlnLysGlyLysGlyAlaIaLysPro | 270 |
| Db | 70777 | -----CCGAGTCTGATGACTGGCAAAAAAGCAACAMAGATATTAACCC | 70733 |
| QY | 271 | LysMetAsnMetLysAspSerPheLeuTyrLeuAspArgSerProTyrIleLeuLeuLeu | 290 |
| Db | 70732 | GTCCAGAACAAATGGTGAT--ATTCTGTACAGAGAAAGAAATTAATATTAATATCTT | 70678 |
| QY | 291 | -----ThrLeuLeuValIleAlaTyrGlyIleCysIleAsnLeu | 303 |
| Db | 70675 | CTGCTCTGTATGCAATTGTCATGCTCTATATATGTTTTCAGGGCTACTCGCAAAACGT | 70616 |
| QY | 304 | IleGluValThrTrpLysSerGln-----LeuLysLeuGlnTyrPro | 317 |
| Db | 70615 | GCATTATATCAAGATTATGGCATTTGCTGTGCGGCTGTTTATATCATGTTTATATATCAG | 70556 |
| QY | 318 | AsnMetAsnAspLysSerGluPheMetGlyAsnPheSerPheTyr---ThrGlyValVal | 336 |
| Db | 70555 | GGATATGGG-----AAAAGATGGCAACAGCATATAG | 70523 |
| QY | 337 | SerValLeuIleMetLeuPheValGlyIaAsnValIleArgLysPheGlyTyr--Leu | 355 |
| Db | 70522 | TTTGCTCTGCTGGTTATGTTTC-----TGTTTCCTTCACGATGGCCACATG | 70478 |
| QY | 356 | ThrGlyAlaLeuValThrProValMetValLeuLeu-----ThrGlyIle | 370 |
| Db | 70477 | CAGGCTTTTCTTCTACATAGCTTAAGTTGATATGACAGTATTCACCGAAACTGTGGCT | 70418 |
| QY | 371 | ValPhePheAlaLeuValIlePheArgAsnGlnAlaIaSerGlyLeuValAlaMetPhe-- | 389 |
| Db | 70417 | CTCATATTATAGCTGGAGGATTTTGGCTCTGACGCTGGAAGTTGATAGTGGTTTATAG | 70358 |
| QY | 390 | -----GlyIhr---ThrProLeuMetLeuAlaValValValGly-----Ala | 402 |
| Db | 70357 | GGGCACTGCTTGGAAACCGAAAGACATATGTTATATCACTTCTATATCGCTGAGCTTGT | 70298 |

```
Oy 403 IlegInAsnIleuSerIysSerThrIysIYr-----AlaLeuPheAspSer 418
|||
|||
Db 70237 ATTATCCCCGTTTCTTCTTGACCGGCGACTATGATGCTTACTGGCCTTCTTAATCTTT 70238
Oy 419 ThrIysGIuMet-----AlaYrIlePro----- 426
|||
|||
Db 70237 ACTCAGCAGGTTTGTGTCAGGGGATGTGCTCTGTGTCGCGAAATAATATCAGTGTAT 70178
Oy 427 LeuAepGIuInIuIlyValIysGIySaIaIaIleAspValValAlaAlaArgPhe 446
|||
|||
Db 70177 TTTAATGTGACGACGCGCGGAGACTGGGTTTATTATATATGTA----- 70130
Oy 447 GIlySerGIyGIyAlaLeu-----IlegInGIuIyLeuLeuValIleCysGIySer 464
|||
|||
Db 70129 GGTTCACCTGTGTGAGCAGATGCGCCTATCTTGTGCGAGTGTGTGCGTCATACAGT 70070
Oy 465 IlegIyAlaMetThrProTyrLeuAlaValIleLeuLeuPheIleIleAlaIle----- 482
|||
|||
Db 70069 CTTGGTACAGCAATGTGCTCTTGTCTTATCTTAACAATTGTGTCTTGTCTTATT 70010
Oy 483 -----TyrLeu-ValSerAlaThrIyLeuAsnIy 492
|||
|||
Db 70009 GGTTTGATATGCCATCAGCTGTACAGAGATGATGATCCAGAAAGCTGCGCTGAATAT 69950
Oy 492 sleuPheLeuAlaGIuSerAlaLeuIyGIuInIu 504
|||
|||
Db 69949 CATACTGTGATGAGAAACCTTTTATGAGCAGAA 69913
```

Search completed: November 25, 2003, 14:53:12
Job time : 1979 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 09:50:16 ; Search time 28 Seconds
(without alignments)
1768.817 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 515
Sequence: 1 MTKTEKPFGLKRSFLWPIH.....AQSALKEQVAVQSDAPASS 515

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

1: PIR 76: *
2: pir1: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 313 | 60.8 | 515 | 2 | E72089 |
| 2 | 224 | 43.5 | 515 | 2 | E86534 |
| 3 | 33 | 6.4 | 528 | 2 | C71561 |
| 4 | 33 | 6.4 | 529 | 2 | C81714 |
| 5 | 13 | 2.5 | 501 | 2 | B71707 |
| 6 | 13 | 2.5 | 501 | 2 | B97790 |
| 7 | 11 | 2.1 | 589 | 1 | S68205 |
| 8 | 11 | 2.1 | 618 | 2 | G86288 |
| 9 | 11 | 2.1 | 624 | 2 | E96834 |
| 10 | 11 | 2.1 | 631 | 2 | T07420 |
| 11 | 10 | 1.9 | 511 | 2 | B97783 |
| 12 | 10 | 1.9 | 512 | 2 | F71653 |
| 13 | 9 | 1.7 | 463 | 2 | E70657 |
| 14 | 9 | 1.7 | 498 | 1 | UQ0026 |
| 15 | 9 | 1.7 | 498 | 1 | A97710 |
| 16 | 8 | 1.6 | 111 | 2 | D72723 |
| 17 | 8 | 1.6 | 194 | 2 | E64538 |
| 18 | 8 | 1.6 | 197 | 2 | D71970 |
| 19 | 8 | 1.6 | 247 | 2 | A53482 |
| 20 | 8 | 1.6 | 282 | 2 | D72771 |
| 21 | 8 | 1.6 | 307 | 1 | RWMSBC |
| 22 | 8 | 1.6 | 316 | 2 | A57069 |
| 23 | 8 | 1.6 | 357 | 2 | T02246 |
| 24 | 8 | 1.6 | 360 | 2 | D84646 |
| 25 | 8 | 1.6 | 400 | 1 | KIECKA |
| 26 | 8 | 1.6 | 400 | 2 | D91026 |
| 27 | 8 | 1.6 | 400 | 2 | E85870 |
| 28 | 8 | 1.6 | 400 | 2 | AD0313 |
| 29 | 8 | 1.6 | 400 | 2 | AG0798 |

| | | | | | |
|-----|---|-----|------|---|--------|
| 30 | 8 | 1.6 | 472 | 2 | S74886 |
| 31 | 8 | 1.6 | 480 | 2 | AD1932 |
| 32 | 8 | 1.6 | 485 | 1 | ANHU |
| 33 | 8 | 1.6 | 552 | 1 | TVMVT3 |
| 34 | 8 | 1.6 | 1772 | 2 | T36105 |
| 35 | 8 | 1.6 | 1964 | 2 | T09059 |
| 36 | 7 | 1.4 | 80 | 2 | S10464 |
| 37 | 7 | 1.4 | 81 | 1 | T5P2A |
| 38 | 7 | 1.4 | 82 | 2 | UCS892 |
| 39 | 7 | 1.4 | 83 | 1 | NSCH1F |
| 40 | 7 | 1.4 | 105 | 1 | NSCH4 |
| 41 | 7 | 1.4 | 108 | 2 | E90163 |
| 42 | 7 | 1.4 | 110 | 1 | C64430 |
| 43 | 7 | 1.4 | 116 | 2 | G70831 |
| 44 | 7 | 1.4 | 146 | 2 | E54403 |
| 45 | 7 | 1.4 | 146 | 2 | D54403 |
| 46 | 7 | 1.4 | 152 | 2 | H71129 |
| 47 | 7 | 1.4 | 166 | 2 | D86026 |
| 48 | 7 | 1.4 | 172 | 2 | S36007 |
| 49 | 7 | 1.4 | 173 | 2 | T11831 |
| 50 | 7 | 1.4 | 176 | 2 | AB3363 |
| 51 | 7 | 1.4 | 191 | 1 | JH0585 |
| 52 | 7 | 1.4 | 193 | 2 | I50693 |
| 53 | 7 | 1.4 | 220 | 2 | AE3321 |
| 54 | 7 | 1.4 | 235 | 2 | AB3298 |
| 55 | 7 | 1.4 | 239 | 2 | D71856 |
| 56 | 7 | 1.4 | 240 | 2 | A64553 |
| 57 | 7 | 1.4 | 245 | 2 | G82068 |
| 58 | 7 | 1.4 | 252 | 2 | F64934 |
| 59 | 7 | 1.4 | 252 | 2 | H90935 |
| 60 | 7 | 1.4 | 252 | 2 | D85784 |
| 61 | 7 | 1.4 | 259 | 2 | S50337 |
| 62 | 7 | 1.4 | 260 | 2 | AE2840 |
| 63 | 7 | 1.4 | 263 | 2 | G97617 |
| 64 | 7 | 1.4 | 264 | 2 | T47468 |
| 65 | 7 | 1.4 | 264 | 2 | F88504 |
| 66 | 7 | 1.4 | 268 | 2 | H70118 |
| 67 | 7 | 1.4 | 268 | 2 | D97548 |
| 68 | 7 | 1.4 | 272 | 2 | C87711 |
| 69 | 7 | 1.4 | 275 | 2 | AH1435 |
| 70 | 7 | 1.4 | 275 | 2 | A11077 |
| 71 | 7 | 1.4 | 276 | 2 | C75414 |
| 72 | 7 | 1.4 | 277 | 2 | T29611 |
| 73 | 7 | 1.4 | 284 | 2 | C97380 |
| 74 | 7 | 1.4 | 285 | 2 | B69653 |
| 75 | 7 | 1.4 | 296 | 2 | T34256 |
| 76 | 7 | 1.4 | 301 | 2 | T33702 |
| 77 | 7 | 1.4 | 306 | 2 | AG2698 |
| 78 | 7 | 1.4 | 309 | 2 | G97480 |
| 79 | 7 | 1.4 | 309 | 2 | T46463 |
| 80 | 7 | 1.4 | 315 | 2 | AB3435 |
| 81 | 7 | 1.4 | 317 | 2 | F96656 |
| 82 | 7 | 1.4 | 320 | 2 | AD2066 |
| 83 | 7 | 1.4 | 320 | 2 | B97206 |
| 84 | 7 | 1.4 | 322 | 2 | T06595 |
| 85 | 7 | 1.4 | 323 | 2 | H91279 |
| 86 | 7 | 1.4 | 323 | 2 | H86120 |
| 87 | 7 | 1.4 | 323 | 2 | S56457 |
| 88 | 7 | 1.4 | 324 | 2 | T27302 |
| 89 | 7 | 1.4 | 329 | 2 | T16469 |
| 90 | 7 | 1.4 | 330 | 2 | C69329 |
| 91 | 7 | 1.4 | 335 | 2 | H96752 |
| 92 | 7 | 1.4 | 336 | 2 | AB1324 |
| 93 | 7 | 1.4 | 336 | 2 | AD1695 |
| 94 | 7 | 1.4 | 336 | 2 | H82040 |
| 95 | 7 | 1.4 | 346 | 2 | T11325 |
| 96 | 7 | 1.4 | 348 | 2 | F83993 |
| 97 | 7 | 1.4 | 352 | 2 | AG0762 |
| 98 | 7 | 1.4 | 354 | 2 | S70595 |
| 99 | 7 | 1.4 | 364 | 2 | F95177 |
| 100 | 7 | 1.4 | 364 | 2 | H98043 |

phytene dehydroge
hypothetical prote
angiotensin precu
transforming prote
probable large gly
notch4 - mouse
fasciculin-like p
short toxin 2 prec
cobrotoxin b - Ch1
short neurotoxin 1
nonhistone chromos
18U ribosomal prot
ribosomal protein
hypothetical prote
hemoglobin C beta
hypothetical prote
hypothetical prote
NADH2 dehydrogenas
peptidylprolyl iso
ribosomal protein
interferon - chick
conserved hypobet
hypothetical prote
probable holocyto
probable holocyto
ribosomal large su
hypothetical prote
hypothetical prote
hypothetical prote
cytochrome-c oxida
ABC transporter, m
probable ABC trans
hypothetical prote
protein B0361.2 [i
glucosamine-6-phos
lipase-protein 11
hydrolyase, alpha/b
PTS system, mannos
PTS system, mannos
hypothetical prote
hypothetical prote
probable ATP-bind1
transmembrane lipi
hypothetical prote
hypothetical prote
cation efflux syst
hypothetical prote
hypothetical prote
prochome ix farne
cobalamn biosynth
methyl-accepting c
probable cytochrom
probable transport
probable transport
hypothetical 34.0K
hypothetical prote
hypothetical prote
immunogenic protei
hypothetical prote
transcription regu
general secretion
NADH2 dehydrogenas
C4-dicarboxylate t
probable membrane
NADH2 dehydrogenas
transulfuration en
O-succinylhomoseri

ALIGNMENTS

RESULT 1

ADP, ATP carrier protein CP0408 [imported] - Chlamydia pneumoniae (strain CWL029 at E72089
 C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C/Accession: E72089; D81580
 R/Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999
 A/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A/Reference number: A72000; MUID:99206606; PMID:10192388
 A/Accession: E72089
 A/Molecule type: DNA
 A/Residues: 1-515 <ARN>
 A/Cross-references: GB:AE001619; GB:AE001363; NID:g4376620; PIDN:AD18495.1; PID:g437662
 A/Experimental source: strain CWL029
 R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A/Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
 A/Reference number: A81500; MUID:20150255; PMID:10684935
 A/Accession: D81580
 A/Molecule type: DNA
 A/Residues: 1-515 <REA>
 A/Cross-references: GB:AE002202; GB:AE002161; NID:g7189324; PIDN:AA38252.1; PID:g718933
 A/Experimental source: strain AR39, HL cells
 C/Genetics: act 1; CP0408
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query March 60.8%; Score 313; DB 2; Length 515;
 Best Local Similarity 99.6%; Pred. No. 3.5e-302;
 Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTKEKRPFGKLRFLPFIHTEHKKVLPMLFECIFNTVLRDTKTLIVAGPSGA 60
 DB 1 MTKEKRPFGKLRFLPFIHTEHKKVLPMLFECIFNTVLRDTKTLIVAGPSGA 60
 QY 61 EALPFIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPFLIFPALFPIYPLRDVL 120
 DB 61 EALPFIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPFLIFPALFPIYPLRDVL 120
 QY 121 HPTFAPRLQATLPPGLGLVAIIRNMPFAFYVLAELMGVMSLMFWGANTITHE 180
 DB 121 HPTFAPRLQATLPPGLGLVAIIRNMPFAFYVLAELMGVMSLMFWGANTITHE 180
 QY 181 AKRYALFGICANISLASGRAIYASGLRASVEGVDPMGISRLMAMTIVSGLVMA 240
 DB 181 AKRYALFGICANISLASGRAIYASGLRASVEGVDPMGISRLMAMTIVSGLVMA 240
 QY 241 SYWINKNVLTDPFPYNEQKKGAKPRNNKDSFLYIDRSPYITLLTLVIAVIGIC 300
 DB 241 SYWINKNVLTDPFPYNEQKKGAKPRNNKDSFLYIDRSPYITLLTLVIAVIGIC 300
 QY 301 INLEIYWKSQLQKQYPMNDYSEFMGNSFWTGVSVTLMLFVGANVIRKFGWLTALV 360
 DB 301 INLEIYWKSQLQKQYPMNDYSEFMGNSFWTGVSVTLMLFVGANVIRKFGWLTALV 360
 QY 361 TPVWVLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVGAIONILSKTKYALFPSTK 420
 DB 361 TPVWVLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVGAIONILSKTKYALFPSTK 420
 QY 421 EMAYIPLDQKQVKGKAAIDVVAARFGKSGGALLIQGGLVTCGSIQAMTPYLAIVLLFTI 480
 DB 421 EMAYIPLDQKQVKGKAAIDVVAARFGKSGGALLIQGGLVTCGSIQAMTPYLAIVLLFTI 480
 QY 481 AIWIVSATKINKFLAQSALKEQVADSDAPASS 515
 DB 481 AIWIVSATKINKFLAQSALKEQVADSDAPASS 515

RESULT 2

ADP/ATP translocase [imported] - Chlamydia pneumoniae (strain J138) E86534
 C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C/Accession: E86534
 R/Shirai, M.; Hiraoka, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ist Nucleic Acids Res. 28, 2311-2314, 2000
 A/Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
 A/Reference number: A86491; MUID:20330349; PMID:10871362
 A/Accession: E86534
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-515 <STO>
 A/Cross-references: GB:BA000008; NID:g8978723; PIDN:BAA98559.1; GSPDB:GN00142
 A/Experimental source: strain J138
 C/Genetics: act 1
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query March 43.5%; Score 224; DB 2; Length 515;
 Best Local Similarity 100.0%; Pred. No. 7.9e-214;
 Matches 224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 LTVIAYGICINLIEVTKWSQLQYPMNDYSEFMGNSFWTGVSVTLMLFVGANVIRK 351
 DB 292 LTVIAYGICINLIEVTKWSQLQYPMNDYSEFMGNSFWTGVSVTLMLFVGANVIRK 351
 QY 352 FGWLTGALVTPVWVLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVGAIONILSKST 411
 DB 352 FGWLTGALVTPVWVLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVGAIONILSKST 411
 QY 412 KYALPDSKEMAYIPLDQKQVKGKAAIDVVAARFGKSGGALLIQGGLVTCGSIQAMTPY 471
 DB 412 KYALPDSKEMAYIPLDQKQVKGKAAIDVVAARFGKSGGALLIQGGLVTCGSIQAMTPY 471
 QY 472 LAVILFTIWIWVATKINKFLAQSALKEQVADSDAPASS 515
 DB 472 LAVILFTIWIWVATKINKFLAQSALKEQVADSDAPASS 515

RESULT 3

C71561
 probable adp/ATP translocase - Chlamydia trachomatis (serotype D, strain UM3/Cx)
 C/Species: Chlamydia trachomatis
 C/Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 20-Aug-1999
 C/Accession: C71561
 R/Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
 A/Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tract
 A/Reference number: A71570; MUID:99000809; PMID:9784136
 A/Accession: C71561
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-528 <ARN>
 A/Cross-references: GB:AE001281; GB:AE001273; NID:g3328454; PIDN:AA67656.1; PID:g3328454
 A/Experimental source: serotype D, strain UM-3/Cx
 C/Genetics: act 1
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query March 6.4%; Score 33; DB 2; Length 528;
 Best Local Similarity 100.0%; Pred. No. 3.4e-24;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 419 TKEMAYIPLDQKQVKGKAAIDVVAARFGKSGG 451
 DB 417 TKEMAYIPLDQKQVKGKAAIDVVAARFGKSGG 449

RESULT 4

ADP, ATP carrier protein TC0335 [imported] - Chlamydia muridarum (strain N199) C81714

C/Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
 C/Accession: G81714
 R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A/Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A/Reference number: A81500; MUID:20150255; PMID:10684935
 A/Accession: G81714
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-529 <TEN>
 A/Cross-references: GB:AE002301; GB:AE002160; NID:g7190372; PIDN:AAF39198.1; PID:g719037
 A/Experimental source: strain N19g (Mopn)
 C/Genetics:
 A/Gene: TC0335
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 6.4%; Score 33; DB 2; Length 529;
 Best Local Similarity 100.0%; Pred. No. 3,4e-24;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 419 TKEMAYIPLDQEQVKGAIDVVAARFGKSG 451
 Db 417 TKEMAYIPLDQEQVKGAIDVVAARFGKSG 449

RESULT 5
 B71707
 ADP, ATP carrier protein (tlc3) RP477 - Rickettsia prowazekii
 C/Species: Rickettsia prowazekii
 C/Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
 C/Accession: B71707
 R/Anderson, S.G.E.; Zomrodipour, A.; Anderson, J.O.; Scheritz-Ponten, T.; Almark, U Nature 396, 133-140, 1998
 A/Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
 A/Reference number: A71630; MUID:99039499; PMID:9823893
 A/Accession: B71707
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-501 <AND>
 A/Cross-references: GB:AU235271; GB:AU235269; NID:g3868717; PIDN:CAA14932.1; PID:g386103
 A/Experimental source: strain Madrid E
 C/Genetics:
 A/Gene: tlc3; RP477
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 2.5%; Score 13; DB 2; Length 501;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 415 LFDSTKEMAYIPL 427
 Db 415 LFDSTKEMAYIPL 427

RESULT 6
 B97790
 ADP, ATP carrier protein [imported] - Rickettsia conorii (strain Malish 7)
 C/Species: Rickettsia conorii
 C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
 C/Accession: B97790
 R/Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R Science 293, 2093-2098, 2001
 A/Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
 A/Reference number: A97700; MUID:21442074; PMID:11557893
 A/Accession: B97790
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-501 <KRB>
 A/Cross-references: GB:AE006914; PIDN:AAU03260.1; PID:g15619815; GSPDB:GN00173
 C/Genetics:
 A/Gene: tlc3

C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 2.5%; Score 13; DB 2; Length 501;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 415 LFDSTKEMAYIPL 427
 Db 415 LFDSTKEMAYIPL 427

RESULT 7
 S68205
 ATP/ADP translocase ATP precursor - Arabidopsis thaliana
 N/Alternate names: adenine nucleotide translocase; ATP/ADP translocator protein
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 22-Jun-1999
 C/Accession: S68205
 R/Kampfenkel, K.; Mochmann, T.; Batz, O.; van Montagu, M.; Inze, D.; Neuhaus, H.E. FEBS Lett. 374, 351-355, 1995
 A/Title: Molecular characterization of an Arabidopsis thaliana cDNA encoding a novel put
 A/Reference number: S68205; MUID:96059943; PMID:7589569
 A/Accession: S68205
 A/Molecule type: mRNA
 A/Residues: 1-589 <KAM>
 A/Cross-references: EMBL:Z49227; NID:g1051108; PIDN:CAA89201.1; PID:g1051109
 C/Genetics:
 A/Gene: ATP1
 A/Genome: nuclear
 C/Superfamily: rickettsial-type ATP/ADP translocase
 F/Keywords: chloroplast; transmembrane protein
 F/100/Domain: transit peptide (chloroplast) #status predicted <TNP>
 F/101-589/Product: ATP/ADP translocase ATP1 #status predicted <MAT>

Query Match 2.1%; Score 11; DB 1; Length 589;
 Best Local Similarity 100.0%; Pred. No. 0.027;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 432 KVGKAIDIV 442
 Db 509 KVGKAIDIV 519

RESULT 8
 G86288
 Probable adenine nucleotide translocase [imported] - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
 C/Accession: G86288
 R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Hu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hutzar, L. Nature 408, 816-820, 2000
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H. A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, I. ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W. A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: A86141; MUID:21016719; PMID:11130712
 A/Accession: G86288
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-618 <STO>
 A/Cross-references: GB:AE005172; NID:g8072388; PIDN:AAF71976.1; GSPDB:GN00141
 C/Genetics:
 A/Map position: 1
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 2.1%; Score 11; DB 2; Length 618;
 Best Local Similarity 100.0%; Pred. No. 0.027;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 432 KVKGAALDVV 442
|||||
Db 509 KVKGAALDVV 519

RESULT 9

adenine nucleotide translocase, 19474-21800 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
C/Accession: E96834
R/Theologian: A.; Ecker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.P.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maltl, R.; Marzalli,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A/Reference number: A6141; MUID:21016719; PMID:11130712
A/Accession: E96834
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-624 <STO>
A/Cross-references: GB:AE005173; NID:96751705; PIDN:AA27687.1; GSPDB:GN00141
C/Genetics:
A/Gene: F516.5
A/Map position: 1
C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 2.1%; Score 11; DB 2; Length 624;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 432 KVKGAALDVV 442
|||||
Db 512 KVKGAALDVV 522

RESULT 10

ATP/ADP-transporter, chloroplast - potato
C/Species: Solanum tuberosum (potato)
C/Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 20-Jun-2000
C/Accession: T07420
R/Tjaden, J.; Mohlmann, T.; Kampfenkel, K.; Heinrichs, G.; Neuhaus, H.E.
Plant J. 16, 531-540, 1998
A/Title: Altered plastidic ATP/ADP-transporter activity influences potato (Solanum tuberosum) growth.
A/Reference number: Z16025
A/Accession: T07420
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-631 <TJA>
A/Cross-references: EMBL:Y10821; PIDN:CAA71785.1
A/Experimental source: cv. Desiree
C/Genetics:
A/Genome: nuclear
C/Superfamily: rickettsial-type ATP/ADP translocase
C/Keywords: chloroplast; membrane protein

Query Match 2.1%; Score 11; DB 2; Length 631;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 432 KVKGAALDVV 442
|||||
Db 510 KVKGAALDVV 520

RESULT 11

B97783

ADP/ATP carrier protein [imported] - Rickettsia conorii (strain Malish 7)
C/Species: Rickettsia conorii
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C/Accession: B97783
R/Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
Science 293, 2093-2098, 2001
A/Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A/Reference number: A97700; MUID:21442074; PMID:11557693
A/Accession: B97783
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-511 <KOR>
A/Cross-references: GB:AE006914; PIDN:ALU03204.1; PID:G15619755; GSPDB:GN00173
C/Genetics:
A/Gene: tlc4
C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 1.9%; Score 10; DB 2; Length 511;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 415 LFDSTKEMAY 424
|||||
Db 422 LFDSTKEMAY 431

RESULT 12

ADP, ATP carrier protein (tlc4) RP500 - Rickettsia prowazekii
C/Species: Rickettsia prowazekii
C/Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C/Accession: F71653
R/Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Pentzen, T.; Alsmark, U
Nature 396, 133-140, 1998
A/Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A/Reference number: A71630; MUID:99039499; PMID:9823893
A/Accession: F71653
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-512 <AND>
A/Cross-references: GB:AJ235272; GB:AJ235269; NID:93861033; PIDN:CAA14952.1; PID:G386105
A/Experimental source: strain Madrid E
C/Genetics:
A/Gene: tlc4; RP500
C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 1.9%; Score 10; DB 2; Length 512;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 415 LFDSTKEMAY 424
|||||
Db 422 LFDSTKEMAY 431

RESULT 13

hypothetical protein Rv2529 - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C/Accession: E70657
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: E70657
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-463 <COL>
A/Cross-references: GB:Z83863; GB:AL123456; NID:93261685; PIDN:CAB06179.1; PID:G1781073

A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV2529
C:Superfamily: Mycobacterium tuberculosis hypothetical protein RV2529

Query Match
Best Local Similarity 100.0%; Pred. No. 2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 AADVVAAAR 445
DB 218 AADVVAAAR 226

RESULT 14
UQ0026
ATP/ADP translocase tlc1 - Rickettsia prowazekii
N:Alternate names: ADP/ATP carrier protein tlc1; RP053
C:Species: Rickettsia prowazekii
C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 03-Nov-2000
C:Accession: JQ0026; E71713
R:Williamson, L.R.; Plano, G.V.; Winkler, H.H.; Krause, D.C.; Wood, D.O.
Gene 80, 269-278, 1989
A:Title: Nucleotide sequence of the Rickettsia prowazekii ATP/ADP translocase-encoding gene
A:Reference number: JQ0026; MUID:9060776; PMID:2555259
A:Accession: JQ0026
A:Molecule type: DNA
A:Residues: 1-498 <MIL>
A:Cross-references: GB:M28916; NID:g152469; PIDN:AAA26382.1; PID:g152470
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Slicheritz-Ponten, T.; Alsmark, U.
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:9823893
A:Accession: E71713
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-498 <AND>
A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14524.1; PID:g386062
A:Experimental source: strain Madrid E
C:Comment: This ATP/ADP translocase shares no detectable amino acid sequence homologies
C:Genetics:
A:Gene: tlc1; RP053
C:Function:
A:Description: exchanges ADP in the rickettsial cell with ATP in the host cell
C:Superfamily: rickettsial-type ATP/ADP translocase
C:Keywords: transmembrane protein
F:28-45/Domain: transmembrane #status predicted <TR01>
F:68-84/Domain: transmembrane #status predicted <TR02>
F:93-115/Domain: transmembrane #status predicted <TR03>
F:183-206/Domain: transmembrane #status predicted <TR04>
F:219-237/Domain: transmembrane #status predicted <TR05>
F:279-297/Domain: transmembrane #status predicted <TR06>
F:346-370/Domain: transmembrane #status predicted <TR07>
F:380-399/Domain: transmembrane #status predicted <TR08>
F:445-461/Domain: transmembrane #status predicted <TR09>
F:466-482/Domain: transmembrane #status predicted <TR10>

Query Match
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 NKXVLTDP 254
DB 241 NKXVLTDP 249

RESULT 15
A97710
ADP/ATP carrier protein [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: A97710
R:Gogate, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barde, V.; Samson, D.; Rd

Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: A97710
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-498 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL02619.1; PID:g15619119; GSPDB:GN00173
C:Genetics:
A:Gene: tlc1
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 NKXVLTDP 254
DB 241 NKXVLTDP 249

RESULT 16
D72723
hypothetical protein APE0325 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: D72723
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahara, H.; Takamizawa, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kikawa, R.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum
A:Reference number: A72450; MUID:99310359; PMID:10382966
A:Accession: D72723
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-111 <KAM>
A:Cross-references: DDBJ:AP000059; NID:g5103911; PIDN:BAA79280.1; PID:d1043066; PID:g5103
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0325

Query Match
Best Local Similarity 100.0%; Pred. No. 6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 LAVVGAI 403
DB 3 LAVVGAI 10

RESULT 17
E64538
hypothetical protein HP0149 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: E64538
R:Tom, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, L.; Son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: E64538
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-194 <TOM>
A:Cross-references: GB:AE000536; GB:AE000511; NID:g2313230; PIDN:AAD07223.1; PID:g2313240

Query Match
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 286 YILLTLTL 293
 |||||
 Db 175 YILLTLTL 182

RESULT 18
 D1970
 hypochlorite protein jhp0137 - Helicobacter pylori (strain J99)
 C/Species: Helicobacter pylori
 A/Variety: strain J99
 C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
 C/Accession: D1970
 R/Alt: R.A.; Ling, L.S.L.; Mofit, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 ; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
 Nature 397, 176-180, 1999
 A/Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A/Reference number: A71800; MUID:99120557; PMID:9923682
 A/Accession: D1970
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-197 <ARN>
 A/Cross-references: GB:AE001452; GB:AE001439; NID:94154639; PIDN:AD05716.1; PID:9415464
 A/Experimental source: strain J99
 C/Genetics:
 A/Gene: jhp0137

Query Match 1.6%; Score 8; DB 2; Length 197;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 286 YILLTLTL 293
 |||||
 Db 178 YILLTLTL 185

RESULT 19
 AS3482
 transmembrane envelope protein Env - primate T-cell lymphotropic virus type L (fragment)
 C/Species: primate T-cell lymphotropic virus type L, PTIV-L
 C/Date: 06-Oct-1994 #sequence_revision 23-Apr-1997 #text_change 19-May-2000
 C/Accession: AS3482
 R/Goubau, P.; Van Brussel, M.; Vandamme, A.M.; Liu, H.F.; Desmyter, J.
 Proc. Natl. Acad. Sci. U.S.A. 91, 2848-2852, 1994
 A/Title: A primate T-lymphotropic virus, PTIV-L, different from human T-lymphotropic vir
 A/Reference number: AS3482; MUID:94195839; PMID:7908445
 A/Accession: AS3482
 A/Status: preliminary
 A/Molecule type: genomic RNA
 A/Residues: 1-247 <GCU>
 A/Cross-references: GB:Z29673; NID:94747831; PIDN:CA82771.1; PID:9483564
 A/Note: sequence extracted from NCBI backbone (NCBI:146143, NCBI:146146)
 A/Note: isolate STIV-PH969; host Papio hamadryas
 C/Superfamily: type C retrovirus env polyprotein

Query Match 1.6%; Score 8; DB 2; Length 247;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 480 IAIWLVA 487
 |||||
 Db 73 IAIWLVA 80

RESULT 20
 D72771
 probable bacteriophage phi11 synthase APE0159 - Aeropyrum pernix (strain K1)
 C/Species: Aeropyrum pernix
 C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C/Accession: D72771
 R/Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
 awa, H.; Takamuta, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kidoh, Y.; Yamazaki, J.; K
 DNA Res. 6, 83-101, 1999
 A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr

A/Reference number: A72450; MUID:99310339; PMID:10382966
 A/Accession: D72771
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-282 <KAM>
 A/Cross-references: DDBJ:AP000058; NID:95103388; PIDN:BAA79070.1; PID:95103549
 A/Experimental source: strain K1
 C/Genetics:
 A/Gene: APE0159
 C/Superfamily: Methanococcus jannaschii conserved hypothetical protein M0279

Query Match 1.6%; Score 8; DB 2; Length 282;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 287 ILLTLTLV 294
 |||||
 Db 234 ILLTLTLV 241

RESULT 21
 RWSBC
 T-cell receptor beta chain precursor (F5) - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 03-Aug-1984 #sequence_revision 01-Dec-2000 #text_change 01-Dec-2000
 C/Accession: S03716; S25118; A02134; A93335; B93333
 R/Palmer, M.S.; Bentley, A.; Gould, K.; Townsend, A.R.M.
 Nucleic Acids Res. 17, 2353, 1989
 A/Title: The T cell receptor from an influenza-A specific murine CTL clone.
 A/Reference number: S03715; MUID:89202046; PMID:2784852
 A/Accession: S03716
 A/Molecule type: mRNA
 A/Residues: 1-307 <PAL>
 A/Cross-references: EMBL:X14388; NID:954668; PIDN:CA32563.1; PID:954669
 R/Austrup, F.; Kodella, V.; Kucharz, T.; Kisch, E.
 submitted to the EMBL Data Library, July 1992
 A/Description: Characterization of idiotypic-specific I-Eg-restricted T suppressor lymphoc
 /c mice.
 A/Reference number: S25117
 A/Accession: S25118
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-'G'GT-'6','C','8','V'FV-'14','I','16','N'-'19','D','21-23','T','25-29','E','31','AE','3'
 ;'K','86','E','88','L','90','S','92','F','94-99','P','101','E','103','K','105-114','LDNSAE'-'17','125-126'
 A/Cross-references: EMBL:X67128; NID:954678; PIDN:CA47607.1; PID:954679
 C/Superfamily: immunoglobulin C region; immunoglobulin homology
 C/Keywords: glycoprotein; heterodimer; receptor; T-cell; transmembrane protein
 F/1-19/Domain: signal sequence #status predicted <SIG>
 F/20-307/Product: T-cell receptor beta chain #status predicted <MAT>
 F/188-228/Domain: immunoglobulin homology <IMM>
 F/281-302/Domain: transmembrane #status predicted <TM>
 F/303-307/Domain: intracellular #status predicted <INT>

Query Match 1.6%; Score 8; DB 1; Length 307;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 233 VSGIVLMA 240
 |||||
 Db 293 VSGIVLMA 300

RESULT 22
 A57069
 olfactory receptor FAT11 - human
 C/Species: Homo sapiens (man)
 C/Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 26-Aug-1999
 C/Accession: A57069
 R/Fan, W.; Liu, Y.C.; Parimoo, S.; Weissman, S.M.
 Genomics 27, 119-123, 1995
 A/Title: Olfactory receptor-like genes are located in the human major histocompatibility
 A/Reference number: A57069; MUID:95394447; PMID:765158
 A/Accession: A57069

A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-316 <FNA>
 A;Cross-references: GB:J35475; NID:g1041044; PIND:AA36567.1; PID:g601919
 C;Genetics:
 A;Gene: GDB:FAT11; OLF2
 A;Cross-references: GDB:1323249; OMIM:600578
 A;Map position: 6p21.3-6p21.3
 C;Superfamily: olfactory receptor OR14

Query Match 1.6%; Score 8; DB 2; Length 316;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
 DB 274 LFYAVGTP 281

RESULT 23
 T02246
 hypothetical protein P1.11659_4 - human
 C;Species: Homo sapiens (man)
 C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 11-Jan-2000
 C;Accession: T02246

R;Lamerdin, J.E.; McCreedy, P.M.; Skowronski, E.; Adamson, A.W.; Burkhardt-Schultz, K.; Christensen, M.; Georgescu, A.; Avila, J.; Liu, S.; Attix, C.; Andreise, T.; Trankiel submitted to the EMBL Data Library, March 1998
 A;Authors: Duarte, S.; Lucas, S.; Bruce, R.; Thomas, P.; Quan, G.; Kronmiller, B.; Arellano, A.; Description: Sequence analysis of a human P1 clone containing the XRC9 DNA repair gene
 A;Reference number: 214637
 A;Accession: T02246
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-357 <LMA>
 A;Cross-references: EMBL:AC004472; NID:g2984582; PIND:AA07983.1; PID:g2984585
 C;Genetics:
 A;Map position: 9
 A;Intons: 2/3; 48/3; 82/1; 101/3; 135/3; 194/3; 243/1; 269/3; 312/3
 C;Superfamily: erythrocyte band 7 integral membrane protein

Query Match 1.6%; Score 8; DB 2; Length 357;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 SLAASGRA 202
 DB 4 SLAASGRA 11

RESULT 24
 D84646
 hypothetical protein At2g25270 [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C;Accession: D84646

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Umayam, L.; Tallon, L.; euse, D.; Nierman, W.C.; White, O.; Bisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A;Reference number: A84420; MUID:20083487; PMID:106117197

A;Accession: D84646
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-360 <STO>
 A;Cross-references: GB:AE002093; NID:g4567250; PIND:AA023664.1; GSPDB:GN00139
 C;Genetics:
 A;Gene: At2g25270
 A;Map position: 2

Query Match 1.6%; Score 8; DB 2; Length 360;
 Best Local Similarity 100.0%; Pred. No. 16;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 162 VMLSLMFV 169
 DB 325 VMLSLMFV 332

RESULT 25

KIECA
 acetate kinase (EC 2.7.2.1) - Escherichia coli (strain K-12)
 N;Alternate names: acetokinase
 C;Species: Escherichia coli
 C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 01-Mar-2002
 C;Accession: J00498; J0358; F65001
 R;Matsuyama, A.; Yamamoto, H.; Nakano, E.
 J. Bacteriol. 171, 577-580, 1989
 A;Title: Cloning, expression, and nucleotide sequence of the Escherichia coli K-12 ackA gene
 A;Reference number: J00498; MUID:89123075; PMID:2536666
 A;Accession: J00498
 A;Molecule type: DNA

A;Residues: 1-400 <MAT>
 A;Cross-references: GB:W22956; NID:g145170; PIND:AAA23406.1; PID:g145171
 A;Experimental source: strain K12
 A;Note: parts of the sequence were confirmed by protein sequencing
 R;Kakuda, H.; Hosono, K.; Shiotsuki, K.; Ichihara, S.
 J. Biochem. 116, 916-922, 1994
 A;Title: Identification and characterization of the ackA (acetate kinase A)-pta (phosphotransferase) operon.

A;Reference number: J00357; MUID:95189796; PMID:7883769
 A;Accession: J00358
 A;Molecule type: DNA
 A;Residues: 376-400 <KAK>
 A;Cross-references: DDBJ:D17576
 R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Colwell, R.; Rose, D.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: F65001
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA
 A;Residues: 1-400 <BLAT>
 A;Cross-references: GB:AE000318; GB:U00096; NID:g1788623; PIND:AA075356.1; PID:g1788633;
 A;Experimental source: strain K-12, substrain MG1655
 C;Comment: Acetate kinase catalyzes the conversion of acetylphosphate to acetate while ATP is converted to ADP. During anaerobic growth of the organism, this enzyme is also involved in the synthesis of acetate.
 C;Genetics:
 A;Gene: ackA
 C;Superfamily: acetate kinase
 C;Keywords: phosphotransferase

Query Match 1.6%; Score 8; DB 1; Length 400;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 AARFGKSG 450
 DB 363 AARFGKSG 370

RESULT 26

D91026
 acetate kinase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
 C;Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C;Accession: D91026
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gaevari, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome comparison with other E. coli strains.
 A;Reference number: A99629; MUID:21156231; PMID:11258796
 A;Accession: D91026
 A;Status: preliminary

A:Molecule type: DNA
A:Residues: 1-400 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA036603.1; PID:G13362650; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: EC63180
C:Superfamily: acetate kinase

Query Match 1.6%; Score 8; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 443 AARFGKSG 450
|||||
Db 363 AARFGKSG 370

RESULT 27
E85870
acetate kinase (imported) - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: E85870
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, B.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85870
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-400 <STD>
A:Cross-references: GB:AA005174; NID:G12516649; PID:AA057425.1; GSPDB:GN00145; UMGF:Z35
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ackA
C:Superfamily: acetate kinase

Query Match 1.6%; Score 8; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 443 AARFGKSG 450
|||||
Db 363 AARFGKSG 370

RESULT 28
AD0313
acetate kinase (EC 2.7.2.1) (imported) - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001

C:Accession: AD0313
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AD0313
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-400 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC91368.1; PID:G15980557; GSPDB:GN00175
C:Genetics:
A:Gene: ackA
C:Superfamily: acetate kinase
C:Keywords: phosphotransferase

Query Match 1.6%; Score 8; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 443 AARFGKSG 450
|||||
Db 363 AARFGKSG 370

RESULT 29
AG0798
acetate kinase (EC 2.7.2.1) (imported) - Salmonella enterica subsp. enterica serovar Typh
C:Species: Salmonella enterica subsp. enterica serovar Typh
A>Note: This species has also been called Salmonella typh
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AG0798
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,
th, T.; Connerthon, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AG0798

A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-400 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07569.1; PID:G16503561; GSPDB:GN00176
C:Genetics:
A:Gene: STY2567
C:Superfamily: acetate kinase
C:Keywords: phosphotransferase

Query Match 1.6%; Score 8; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 443 AARFGKSG 450
|||||
Db 363 AARFGKSG 370

RESULT 30
S74886
phycoene dehydrogenase (EC 1.3.-.-) - Synechocystis sp. (strain PCC 6803)
N:Alternate names: phycoene desaturase; protein slr1254
C:Species: Synechocystis sp.

A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 04-Feb-2000
C:Accession: S74886; S20939
R:Kaneko, T.; Sato, S.; Kocani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

S.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74886

A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-472 <XNA>

A:Cross-references: EMBL:DB0909; GB:AB00133; NID:G1652844; PID:BA17847.1; PID:dl01858
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

R:Martinez-Ferez, I.M.; Vioque, A.
Plant Mol. Biol. 18, 981-983, 1992
A:Title: Nucleotide sequence of the phycoene desaturase gene from Synechocystis sp. PCC
A:Reference number: S20939; MUID:92256820; PMID:1581575
A:Accession: S20939

A:Molecule type: DNA

A:Residues: 1-451, 'RSPPIST', '459', 'KPF', '463-464', 'PGK' <MAR>
A:Cross-references: EMBL:X62574

C:Genetics:

A:Gene: crtD; pds

C:Keywords: carotenoid biosynthesis; oxidoreductase

Query Match 1.6%; Score 8; DB 2; Length 472;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 139 GLVALLRN 146
|||||||
Db 109 GLVALLRN 116

RESULT 31

AD1932
hypothetical protein all1007 [imported] - Nostoc sp. (strain PCC 7120)

C/Species: Nostoc sp. PCC 7120
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C/Accession: AD1932
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iriuchika, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A/Reference number: AB1807; MUID:21595285; PMID:11759840
A/Accession: AD1932
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-480 <KTR>
A/Cross-references: GB:BA000019; PIDN:BA072964.1; PID:G17130353; GSPDB:GN00179
A/Experimental source: strain PCC 7120
C/Genetics:
A/Gene: all1007

Query Match 1.6%; Score 8; DB 2; Length 480;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 288 LLTLTLVI 295
|||||||
Db 429 LLTLTLVI 436

RESULT 32

ANHU
angiotensin precursor [validated] - human

N/Alternate names: angiotensinogen
N/Contents: angiotensin I; angiotensin II; angiotensin III
C/Species: Homo sapiens (man)
C/Date: 06-Jul-1992 #sequence_revision 19-Jan-1996 #text_change 08-Dec-2000
C/Accession: A35203; A31362; I37168; A30825; I39462; A90487; A90226; I54281; A01
R/Fukami, A.; Takahashi, S.; Seo, M.S.; Tada, M.; Tanimoto, K.; Uehara, S.; Murakami,
U. Biol. Chem. 265, 7576-7582, 1990
A/Title: Structure and expression of the human angiotensinogen gene. Identification of a
A/Accession: A35203
A/Molecule type: DNA
A/Residues: 1-485 <PRK>
A/Cross-references: GB:X15323; GB:X15324; GB:X15325; GB:X15326; GB:X15327
R/Galliard, I.; Clauser, E.; Corvol, P.
DNA 8, 87-99, 1989
A/Title: Structure of human angiotensinogen gene.
A/Reference number: A31362; MUID:89170129; PMID:2924688
A/Accession: A31362
A/Molecule type: DNA
A/Residues: 1-267, 'W', 269-332, 'E', 334-485 <GAI>
A/Cross-references: GB:M24686; GB:M24687; GB:M24688
A/Note: the authors translated the codon GAA for residue 333 as Gln
R.Nibu, Y.; Takahashi, S.; Tanimoto, K.; Murakami, K.; Fukamizu, A.
J. Biol. Chem. 269, 28598-28605, 1994
A/Title: Identification of cell type-dependent enhancer core element located in the 3'-d
A/Reference number: I37168; MUID:95050659; PMID:7961807
A/Accession: I37168
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-285 <NIB1>
A/Cross-references: EMBL:X15324; NID:G1197496; PIDN:CAA33385.1; PID:G1197497
A/Accession: I37169
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA

A/Residues: 287-375 <NIB2>
A/Cross-references: EMBL:X15325; NID:G28695
R/Kunapuli, S.P.; Benedict, C.R.; Kumar, A.
Arch. Biochem. Biophys. 254, 642-646, 1987
A/Title: Tissue specific hormonal regulation of the rat angiotensinogen gene expression.
A/Reference number: A60825; MUID:87212053; PMID:3579322
A/Accession: A60825
A/Molecule type: mRNA

A/Residues: 32-184 <KUN1>
R/Kunapuli, S.P.; Kumar, A.
Circ. Res. 60, 786-790, 1987
A/Title: Molecular cloning of human angiotensinogen cDNA and evidence for the presence o
A/Reference number: I39462; MUID:87244745; PMID:2885106
A/Accession: I39462
A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA
A/Residues: 1-267, 'W', 269-338 <KUN2>
A/Cross-references: GB:M69110; NID:G178643; PIDN:AAA52282.1; PID:G553181
R/Kageyama, R.; Ohkubo, H.; Nakaneishi, S.
Biochemistry 23, 3603-3609, 1984
A/Title: Primary structure of human preangiotensinogen deduced from the cloned cDNA sequ
A/Reference number: A90487; MUID:8500455; PMID:6089875
A/Accession: A90487

A/Molecule type: mRNA
A/Residues: 1-267, 'W', 269-485 <KAG>
A/Cross-references: GB:K02215; NID:G178639; PIDN:AAA51731.1; PID:G178640
A/Note: it is uncertain whether Met-1 or Met-10 is the initiator
R/Tewksbury, D.A.; Dart, R.A.; Travis, J.
Biochem. Biophys. Res. Commun. 99, 1311-1315, 1981
A/Title: The amino terminal amino acid sequence of human angiotensinogen.
A/Reference number: A90226; MUID:81255848; PMID:7259779
A/Accession: A90226

A/Molecule type: protein
A/Residues: 34-46, 'X', 48-50, 'S', 52-57, 'D' <TEW>
R/Hixson, J.E.; Powers, P.K.
Hum. Genet. 96, 110-112, 1995
A/Title: Detection and characterization of new mutations in the human angiotensinogen ge
A/Reference number: I54281; MUID:95331754; PMID:7607642
A/Accession: I54281
A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA
A/Residues: 176-241, 'I', 243-267, 'W', 269-287, 'ANLSAG' <HIX>
A/Cross-references: GB:G78529; NID:G999316; PIDN:AD14287.1; PID:G4261987
C/Comment: Angiotensin I is released from angiotensinogen by renin, which is secreted in
e I (angiotensin-converting enzyme), primarily in the lungs.
C/Comment: The release of the amino-terminal residue (Asp-34) from angiotensin I and ang
sp-1]angiotensin I is converted to angiotensin III by dipeptidyl carboxypeptidase I.
C/Comment: Angiotensin II causes vasoconstriction by direct action on blood vessels, by
o induces thirst.

C/Comment: Angiotensin II and angiotensin III are equally potent in stimulating the synt
C/Comment: Angiotensinogen is synthesized in the liver and secreted into the plasma. The
C/Genetics:
A/Gene: GDB:AGT
A/Cross-references: GDB:118750; OMIM:106150
A/Map position: 1q42-1q43
A/Introns: 286/1; 375/2; 423/3

C/Superfamily: antithrombin III
C/Keywords: blood pressure control; glycoprotein; liver; plasma; vasoconstrictor
F/1-33/Domain: (or 10-33) signal sequence #status predicted <SIG>
F/34-485/Product: angiotensinogen #status predicted <MPT>
F/34-43/Product: angiotensin I #status experimental <PP1>
F/35-41/Product: angiotensin II #status experimental <PP2>
F/47-170,304,328/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 1.6%; Score 8; DB 1; Length 485;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 126 ADRLQAIL 133
|||||||
Db 155 ADRLQAIL 162

RESULT 33

Transforming protein int-3 - mouse mammary tumor virus

C/Species: mouse mammary tumor virus, MMTV

C/Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 10-Jul-1998

C/Accession: A38072

R/Robbins, J.; Blondel, B.J.; Gallahan, D.; Callahan, R.

J. Virol. 66, 2594-2599, 1992

A/Title: Mouse mammary tumor gene int-3: a member of the notch gene family transforms ma

A/Reference number: A38072; PMID:92194507; PMID:1312643

A/Accession: A38072

A/Molecule type: mRNA

A/Residues: 1-552 <ROB>

A/Cross-references: GB:M80456

C/Genetics:

A/Gene: int-3

C/Superfamily: int-3 transforming protein; ankyrin repeat homology

C/Keywords: duplication; oncogene; transforming protein

F/217-248/Domain: ankyrin repeat homology <AN1>

F/249-281/Domain: ankyrin repeat homology <AN2>

F/283-315/Domain: ankyrin repeat homology <AN3>

F/316-348/Domain: ankyrin repeat homology <AN4>

Query Match 1.6%; Score 8; DB 1; Length 552;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 390 GTTPLMLA 397

Db 284 GTTPLMLA 291

RESULT 34

Probable large glycine/alanine rich protein - Streptomyces coelicolor

C/Species: Streptomyces coelicolor

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C/Accession: T36105

R/Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrett, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, April 1999

A/Reference number: Z21597

A/Accession: T36105

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1772 <MUR>

A/Cross-references: EMBL:AL049707; PIDN:CA841271.1; GSPDB:GN00070; SCOEDB:SCEL5.02C

A/Experimental source: strain A3(2)

C/Genetics:

A/Gene: SCOEDB:SCEL5.02C

Query Match 1.6%; Score 8; DB 2; Length 1772;

Best Local Similarity 100.0%; Pred. No. 62;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 GAGSGAG 61

Db 1176 GAGSGAG 1183

RESULT 35

Notch4 - mouse

C/Species: Mus musculus (house mouse)

C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 08-Sep-2002

C/Accession: T09059

R/Bowen, L.; Mahitras, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.; Sc

submitted to the EMBL Data Library, October 1997

A/Reference number: Z16543

A/Accession: T09059

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1964 <ROM>

A/Cross-references: EMBL:AF030001; NID:92564945; PID:92564947

C/Genetics:

A/Gene: notch4

A/Map position: 17

A/Introns: 22/1; 49/2; 148/1; 264/1; 305/1; 384/1; 436/1; 501/1; 539/1; 577/1; 618/1; 67

1679/3; 1729/1; 1761/3

C/Superfamily: notch protein; ankyrin repeat homology; EGF homology

C/Keywords: receptor; signal transduction

F/514-545/Domain: EGF homology <EGF>

Query Match 1.6%; Score 8; DB 2; Length 1964;

Best Local Similarity 100.0%; Pred. No. 67;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 390 GTTPLMLA 397

Db 1696 GTTPLMLA 1703

RESULT 36

fascicullin-like protein - eastern green mamba

N/Alternate names: protein C1381C1

C/Species: Dendroaspis angusticeps (eastern green mamba)

C/Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999

C/Accession: S10464; S21298

R/Ducancel, F.

submitted to the EMBL Data Library, March 1990

A/Reference number: S10464

A/Accession: S10464

A/Molecule type: mRNA

A/Residues: 1-80 <DUC>

A/Cross-references: EMBL:X52260; NID:962757; PIDN:CAA36503.1; PID:962758

R/Ducancel, F.

submitted to the EMBL Data Library, June 1990

A/Reference number: S21298

A/Accession: S21298

A/Molecule type: mRNA

A/Residues: 1-80 <DDW>

A/Cross-references: EMBL:X53410; NID:962753; PIDN:CAA37486.1; PID:962754

C/Superfamily: snake toxin

C/Keywords: toxin

Query Match 1.4%; Score 7; DB 2; Length 80;

Best Local Similarity 100.0%; Pred. No. 45;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 288 LLLTLVY 294

Db 4 LLLTLVY 10

RESULT 37

TSEB2A

short toxin 2 precursor - eastern green mamba

N/Alternate names: protein Daf8; toxin F-VIII; toxin TA2

C/Species: Dendroaspis angusticeps (eastern green mamba)

C/Date: 22-Jun-1981 #sequence_revision 17-Mar-2000 #text_change 17-Mar-2000

C/Accession: S21299; A01675

R/Ducancel, F.

submitted to the EMBL Data Library, June 1990

A/Reference number: S21298

A/Accession: S21299

A/Molecule type: mRNA

A/Residues: 1-81 <DUC>

A/Cross-references: EMBL:X53409; NID:962755; PIDN:CAA37485.1; PID:962756

R/Viljoen, C.C.; Botes, D.P.

J. Biol. Chem. 249, 366-372, 1974

A/Title: Snake venom toxin. The purification and amino acid sequence of toxin TA2 from

A/Reference number: A01675; PMID:74070661; PMID:4203470

A/Accession: A01675

A/Molecule type: protein

A/Residues: 22-53, 'V', 55-81 <VIL>
C/Comment: Short toxin 2 is similar to toxin 1 in its length and in its nonlethal nature
, and thus forms a fifth type of toxin. Its pharmacological action is unknown. It probab
C/Superfamily: snake toxin
C/Keywords: toxin; venom
F/1-21/Domain: signal sequence #status predicted <SIG>
F/22-81/Product: short toxin 2 #status experimental <MAT>
F/24-43,38-60,62-73,74-79/Disulfide bonds: #status predicted

Query Match 1.4%; Score 7; DB 1; Length 81;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 LLLTLV 294
DB 4 LLLTLV 10

RESULT 38

JCS892

cobrotoxin b - Chinese cobra

C/Species: Naja naja atra (Chinese cobra)

C/Date: 18-Mar-1998 #sequence_revision 18-Mar-1998 #text_change 21-Jul-2000

C/Accession: JCS892

R/Chang, L.S.; Chou, Y.C.; Lin, S.R.; Wu, B.N.; Lin, J.; Hong, E.; Sun, Y.J.; Hsiao, C.D

J. Biochem. 122, 1252-1259, 1997

A/Title: A novel neurotoxin, cobrotoxin b, from Naja naja atra (Taiwan cobra) venom: Pur

A/Reference number: JCS892; PMID:98158338; PMID:9458573

A/Accession: JCS892

A/Molecule type: DNA

A/Residues: 1-82 <CHA>

A/Cross-references: GB:Y13399; NID:gs419941; PIDN:CAW3829.2; PID:gs419942

A/Experimental source: venom

C/Genetics:

A/Intons: 19/1; 54/1

C/Superfamily: snake toxin

Query Match 1.4%; Score 7; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 LLLTLV 294
DB 4 LLLTLV 10

RESULT 39

NINJ1F

short neurotoxin 1 precursor [validated] - Chinese cobra

N/Alternate names: cobrotoxin 2

C/Species: Naja naja atra (Chinese cobra)

C/Date: 13-Jul-1981 #sequence_revision 17-Mar-2000 #text_change 15-Sep-2000

C/Accession: JCS769; A01698; D60667

R/Chang, L.; Lin, J.; Chou, Y.; Hong, E.

Biochem. Biophys. Res. Commun. 239, 756-762, 1997

A/Title: Genomic structures of cardiotoxin 4 and cobrotoxin from Naja naja atra (Taiwan

A/Reference number: JCS768; PMID:98042466; PMID:9367842

A/Accession: JCS769

A/Molecule type: DNA

A/Residues: 1-83 <CHA>

A/Cross-references: GB:Y12492; NID:gs5524749; PIDN:CAW3097.2; PID:gs5524750

A/Experimental source: venom

R/Yang, C.C.; Yang, H.J.; Huang, J.S.

Biochim. Biophys. Acta 188, 65-77, 1969

A/Title: The amino acid sequence of cobrotoxin.

A/Reference number: A90575; PMID:59294410; PMID:5820687

A/Accession: A01698

A/Molecule type: protein

A/Residues: 22-83 <YAN>

R/Chou, S.H.; Raynor, R.L.; Zheng, B.; Chambers, T.C.; Kuo, J.F.

Biochemistry 32, 2062-2067, 1993

A/Title: Cobra venom cardiotoxin (cytotoxin) isoforms and neurotoxin: comparative potenc

A/Reference number: A40667; PMID:93192268; PMID:8448165

A/Accession: D40667

A/Molecule type: protein

A/Residues: 22-83 <CHI>

A/Cross-references: PIDN:AB25735.1; PID:gs299271

A/Experimental source: venom

A/Note: sequence extracted from NCBI backbone (NCBIP:127034)

R/Yang, C.C.; Yang, H.J.; Chiu, R.H.C.

Biochim. Biophys. Acta 214, 355-363, 1970

A/Title: The position of disulfide bonds in cobrotoxin.

A/Reference number: A90576; PMID:71125117; PMID:5533659

A/Contents: annotation; disulfide bonds

A/Reference number: A52941; PDB:1COD

A/Contents: annotation; conformation by (1)H-NMR, residues 1-62

R/Yu, C.; Bhaskaran, R.; Chuang, L.C.; Yang, C.C.

Biochemistry 32, 2131-2136, 1993

A/Title: Solution conformation of cobrotoxin: a nuclear magnetic resonance and hybrid di

A/Reference number: A58569; PMID:93183875; PMID:8443154

A/Contents: annotation; conformation by (1)H-NMR

C/Genetics:

A/Intons: 19/1; 55/1

C/Superfamily: snake toxin

C/Keywords: neurotoxin; venom

F/1-21/Domain: signal sequence #status predicted <SIG>

F/22-83/Product: cobrotoxin #status experimental <MAT>

F/24-45,38-62,64-75,76-81/Disulfide bonds: #status experimental

Query Match 1.4%; Score 7; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 LLLTLV 294
DB 4 LLLTLV 10

RESULT 40

NSCH4

nonhistone chromosomal protein HMG-14a - chicken

N/Alternate names: high mobility group protein-14a

C/Species: Gallus gallus (chicken)

C/Date: 30-Jun-1989 #sequence_revision 12-Apr-1996 #text_change 22-Jun-1999

C/Accession: S22122; J0282; J00140

R/Browne, D.L.; Dodgson, J.B.

Submitted to the EMBL Data Library, November 1991

A/Description: The chicken HMG-14a gene is transcribed into multiple mRNAs.

A/Reference number: S22122

A/Accession: S22122

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-105 <BRO>

A/Cross-references: EMBL:X63086

R/Dodgson, J.B.; Browne, D.L.; Black, A.J.

Gene 63, 287-295, 1988

A/Title: Chicken chromosomal protein HMG-14 and HMG-17 cDNA clones: isolation, character

A/Reference number: J0283; PMID:88255874; PMID:3384337

A/Accession: J0282

A/Molecule type: mRNA

A/Residues: 2-105 <DOD>

A/Cross-references: GB:M26675; NID:gs211920; PIDN:AAA48815.1; PID:gs211921

R/Browne, D.L.; Dodgson, J.B.

Gene 124, 199-206, 1993

A/Title: The gene encoding chicken chromosomal protein HMG-14a is transcribed into multi

A/Reference number: J00140; PMID:93185924; PMID:8444343

A/Accession: J00140

A/Molecule type: DNA

A/Residues: 2-5,16-42 <BR2>

A/Cross-references: EMBL:X63083

C/Comment: The HMG proteins are small nonhistone chromosomal proteins, rich in both basic

C/Genetics: This protein is apparently one of the necessary components of actively transcr

A/Reference number: A40667; PMID:93192268; PMID:8448165

A/Reference number: A40667; PMID:93192268; PMID:8448165

A/Reference number: A40667; PMID:93192268; PMID:8448165

A/Reference number: A40667; PMID:93192268; PMID:8448165

A/Reference number: A40667; PMID:93192268; PMID:8448165

A/Reference number: A40667; PMID:93192268; PMID:8448165

A/Reference number: A40667; PMID:93192268; PMID:8448165

A;Introns: 5/3; 15/3; 25/3; 42/3; 90/3
C;Superfamily: nonhistone chromosomal protein HMG-17
C;Keywords: chromosomal protein; DNA binding; nucleus

Query Match 1.4%; Score 7; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 263 KGKKGAK 269
|||||

Db 59 KGKKGAK 65

Search completed: November 25, 2003, 10:07:35
Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 06:57:49 ; Search time 19 Seconds

(without alignments)
1274.671 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 515

Sequence: 1 MTKTEKPKFGLRSLFWPIH.....AQSAKKEGVNQEDSAPASS 515

Scoring table: OLIGO

Searched: 127863 seqs, 47026705 residues

Word size: 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database: SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------|---------------------|
| 1 | 313 | 60.8 | 515 | 1 TLCL1_CHLNP | Q94832 chlamydia p |
| 2 | 333 | 6.4 | 528 | 1 TLCL1_CHLNP | Q94068 chlamydia t |
| 3 | 33 | 6.4 | 529 | 1 TLCL1_CHLNP | Q94068 chlamydia m |
| 4 | 13 | 2.5 | 501 | 1 TLCC1_RICPR | Q94067 rickettsia m |
| 5 | 11 | 2.1 | 618 | 1 TLCC2_ARATH | P92935 rickettsia m |
| 6 | 11 | 2.1 | 624 | 1 TLCL1_ARATH | Q93002 arabidopsis |
| 7 | 11 | 2.1 | 631 | 1 TLCL1_SOLTU | Q24381 solanum tub |
| 8 | 10 | 1.9 | 512 | 1 TLCD1_RICPR | Q94067 rickettsia m |
| 9 | 9 | 1.7 | 498 | 1 TLCA1_RICPR | P19568 rickettsia m |
| 10 | 8 | 1.6 | 173 | 1 TCBB2_MOUSE | P01851 mus musculu |
| 11 | 8 | 1.6 | 312 | 1 OZH2_HUMAN | Q95918 homo sapien |
| 12 | 8 | 1.6 | 316 | 1 OZH1_HUMAN | Q95264 homo sapien |
| 13 | 8 | 1.6 | 316 | 1 OZH3_HUMAN | Q15062 homo sapien |
| 14 | 8 | 1.6 | 400 | 1 ACKA_ECOLI | P15046 escherichia |
| 15 | 8 | 1.6 | 400 | 1 ACKA_SALTY | Q84599 salmonella |
| 16 | 8 | 1.6 | 400 | 1 ACKA_YERPE | Q84599 yersinia pe |
| 17 | 8 | 1.6 | 472 | 1 CRTI_SYNT3 | P29273 synechocyst |
| 18 | 8 | 1.6 | 485 | 1 ANGT_HUMAN | P01019 homo sapien |
| 19 | 8 | 1.6 | 1964 | 1 NTC4_MOUSE | P31695 mus musculu |
| 20 | 8 | 1.6 | 2003 | 1 NTC4_HUMAN | Q99466 homo sapien |
| 21 | 8 | 1.4 | 35 | 1 PETG_CYACA | O99199 cyandium c |
| 22 | 8 | 1.4 | 80 | 1 TX31_DENAN | P18329 dendroaspis |
| 23 | 8 | 1.4 | 81 | 1 TXF8_DENAN | P01404 dendroaspis |
| 24 | 8 | 1.4 | 82 | 1 CBTB_NAUT | P80958 naja atra |
| 25 | 8 | 1.4 | 83 | 1 CBTB_NAUT | P01430 naja atra |
| 26 | 8 | 1.4 | 104 | 1 HG15_CHICK | P12902 gallus gall |
| 27 | 8 | 1.4 | 108 | 1 RL3B_SULSO | Q98063 sulfolobus |
| 28 | 8 | 1.4 | 110 | 1 RL3B_METJA | P54061 methanococ |
| 29 | 8 | 1.4 | 146 | 1 HBBO_PAGRO | P82345 pagotomema |
| 30 | 8 | 1.4 | 146 | 1 HBBO_PAGRO | P45722 pagotomema |
| 31 | 8 | 1.4 | 146 | 1 HBBO_PAGRO | P45722 pagotomema |
| 32 | 8 | 1.4 | 168 | 1 NU6M_LOLAL | O47478 loligo blae |
| 33 | 8 | 1.4 | 172 | 1 NU6M_CYPCA | P24982 cyprinus ca |

| | | | | | |
|-----|---|-----|------|--------------|---------------------|
| 34 | 7 | 1.4 | 173 | 1 NU6M_CARAU | O78689 carassius a |
| 35 | 7 | 1.4 | 173 | 1 NU6M_GADMO | P55783 gadus morhu |
| 36 | 7 | 1.4 | 185 | 1 RRP_WIGER | Q86295 wigleswort |
| 37 | 7 | 1.4 | 191 | 1 RKT4_SPTOL | P27683 spinacia ol |
| 38 | 7 | 1.4 | 192 | 1 INF_MELGA | P51527 melaeagris g |
| 39 | 7 | 1.4 | 193 | 1 INF1_CHICK | P42165 gallus gall |
| 40 | 7 | 1.4 | 193 | 1 INF3_CHICK | Q90872 gallus gall |
| 41 | 7 | 1.4 | 194 | 1 Y537_FUSNN | O84792 fusobacteri |
| 42 | 7 | 1.4 | 236 | 1 YDUX_ECOLI | P76219 escherichia |
| 43 | 7 | 1.4 | 240 | 1 HIS4_OCEBI | Q86851 oceanobacti |
| 44 | 7 | 1.4 | 253 | 1 C151_CERAE | Q94502 ceratophyc |
| 45 | 7 | 1.4 | 253 | 1 C151_HUMAN | P48509 homo sapien |
| 46 | 7 | 1.4 | 264 | 1 YMP2_CABEL | Q10946 caenorhabdi |
| 47 | 7 | 1.4 | 267 | 1 SY72_ARATH | Q94463 arabidopsis |
| 48 | 7 | 1.4 | 268 | 1 NAGE_BORBU | Q30564 borrelia bu |
| 49 | 7 | 1.4 | 277 | 1 SUR4_CABEL | Q18864 caenorhabdi |
| 50 | 7 | 1.4 | 295 | 1 LPLC_BACSU | P39129 bacillus su |
| 51 | 7 | 1.4 | 315 | 1 COXX_BRUME | Q84796 bruceella me |
| 52 | 7 | 1.4 | 322 | 1 CCSA_CYAPA | P48257 cyanophora |
| 53 | 7 | 1.4 | 331 | 1 YJFF_ECOLI | P37772 escherichia |
| 54 | 7 | 1.4 | 336 | 1 GSPK_VIRCH | P45781 vibrio chol |
| 55 | 7 | 1.4 | 340 | 1 ALP_STRGB | Q94507 streptomyce |
| 56 | 7 | 1.4 | 344 | 1 OM40_DROME | Q94416 drosophila |
| 57 | 7 | 1.4 | 354 | 1 NU2M_ASTPE | Q33819 asterina pe |
| 58 | 7 | 1.4 | 375 | 1 GP27_HUMAN | Q94567 homo sapien |
| 59 | 7 | 1.4 | 380 | 1 CYB_IACUV | O48089 lacerta viv |
| 60 | 7 | 1.4 | 380 | 1 CYB_PANRU | Q94567 rana rugosa |
| 61 | 7 | 1.4 | 388 | 1 PEPF_RABIT | P27823 oryctolagus |
| 62 | 7 | 1.4 | 402 | 1 YIN2_STRAM | P32426 streptomyce |
| 63 | 7 | 1.4 | 404 | 1 RSPA_ECOLI | P38104 escherichia |
| 64 | 7 | 1.4 | 442 | 1 YOH2_BACSU | P54505 bacillus su |
| 65 | 7 | 1.4 | 457 | 1 NORM_ECOS7 | P58164 escherichia |
| 66 | 7 | 1.4 | 457 | 1 NORM_ECOLI | P73740 escherichia |
| 67 | 7 | 1.4 | 475 | 1 Z342_HUMAN | Q94504 homo sapien |
| 68 | 7 | 1.4 | 481 | 1 IMDH_HELPJ | Q92114 helicobacte |
| 69 | 7 | 1.4 | 484 | 1 IMDH_HELPJ | P56088 helicobacte |
| 70 | 7 | 1.4 | 484 | 1 Y014_HUMAN | P57764 homo sapien |
| 71 | 7 | 1.4 | 493 | 1 Y014_HUMAN | Q15048 homo sapien |
| 72 | 7 | 1.4 | 516 | 1 CP11_CAVPO | Q06367 cavia porce |
| 73 | 7 | 1.4 | 536 | 1 FLIF_CAUOR | Q04954 caulobacter |
| 74 | 7 | 1.4 | 644 | 1 YNL5_YEAST | P53925 saccharomyc |
| 75 | 7 | 1.4 | 662 | 1 SUT2_STYHA | P53392 stylosanthe |
| 76 | 7 | 1.4 | 667 | 1 SUT1_STYHA | P53391 stylosanthe |
| 77 | 7 | 1.4 | 686 | 1 ATB2_LISIN | Q92637 listeria in |
| 78 | 7 | 1.4 | 686 | 1 KLC_STRPU | Q05090 streptococ |
| 79 | 7 | 1.4 | 783 | 1 YAF3_YEAST | P39719 saccharomyc |
| 80 | 7 | 1.4 | 786 | 1 TLRI_HUMAN | Q15399 homo sapien |
| 81 | 7 | 1.4 | 915 | 1 Y007_CABEL | Q09446 caenorhabdi |
| 82 | 7 | 1.4 | 955 | 1 AGAB_VIRB7 | P48840 vibrio sp. |
| 83 | 7 | 1.4 | 1033 | 1 ITAB_MOUSE | Q94504 mus musculu |
| 84 | 7 | 1.4 | 1137 | 1 BCSC_ECOS7 | Q84504 escherichia |
| 85 | 7 | 1.4 | 1140 | 1 BCSC_ECOLI | P37650 escherichia |
| 86 | 7 | 1.4 | 1255 | 1 VGL2_CVISA | P59594 human coron |
| 87 | 7 | 1.4 | 1363 | 1 XDH_EMENT | Q12553 emeticella |
| 88 | 7 | 1.4 | 1429 | 1 L112_CABEL | P14568 caenorhabdi |
| 89 | 7 | 1.4 | 3011 | 1 POLG_HCVI | P26664 h genome po |
| 90 | 7 | 1.4 | 3011 | 1 POLG_HCVI | P27958 h genome po |
| 91 | 7 | 1.4 | 3678 | 1 DMD_MOUSE | P11531 mus musculu |
| 92 | 7 | 1.4 | 3680 | 1 DMD_CANFA | Q97592 canis famli |
| 93 | 7 | 1.4 | 3685 | 1 DMD_HUMAN | P11533 homo sapien |
| 94 | 6 | 1.2 | 43 | 1 DHBD_ASPOR | P80402 aspergillus |
| 95 | 6 | 1.2 | 46 | 1 RL34_MYCAV | Q92891 mycobacteri |
| 96 | 6 | 1.2 | 48 | 1 OP01_OXKTI | P83247 oxyopes kit |
| 97 | 6 | 1.2 | 52 | 1 IPR_SOLME | P01078 solanum mel |
| 98 | 6 | 1.2 | 54 | 1 LT02_HORVU | Q94504 hordeum vul |
| 99 | 6 | 1.2 | 57 | 1 VSH_MUMPU | P22110 mumps virus |
| 100 | 6 | 1.2 | 58 | 1 COX5_YEAST | P07255 saccharomyc |

RESULT 1

ALIGNMENTS

TLCL CHLPRN STANDARD; PRT; 515 AA.
 AC Q92802; Q92802; 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP carrier protein 1 (ADP/ATP translocase 1).
 GN TLCA OR ADP 1 OR CPN0351 OR CP0408.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.
 NCBI_TaxID=83358;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CM1029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uetzerback T., Berry K., Bass S.,
 RA Libber K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishi K., Hattori M., Kohara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CM1029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 CC EMBL; AE001619; AAD18495.1; -;
 CC EMBL; AE002202; AAF38252.1; -;
 CC EMBL; AP002546; BAA98559.1; -;
 CC PIR; E72089; E82089.
 CC PIR; B86534; B86534.
 CC TIGR; CP0408; -;
 CC InterPro; IPR004667; ADP_ATP-car.
 CC Pfam; PF03219; TLC; 1.
 CC TIGRfam; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 FT TRANSMEM 24 44 POTENTIAL.
 FT TRANSMEM 62 82 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 124 144 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 226 246 POTENTIAL.
 FT TRANSMEM 286 306 POTENTIAL.
 FT TRANSMEM 329 349 POTENTIAL.
 FT TRANSMEM 358 378 POTENTIAL.
 FT TRANSMEM 383 403 POTENTIAL.

FT TRANSMEM 465 485 POTENTIAL.
 FT CONFLICT 141 141 V -> D (IN REF. 3).
 SQ SEQUENCE 515 AA; 57197 MM; F3F4A52C008E106 CRC64;
 Query Match 60.8%; Score 313; DB 1; Length 515;
 Best Local Similarity 99.6%; Pred. No. 2.56-302;
 Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MTKEEKPFGRLRSEFLPIHHEKLVLPMLMFCITENTYVLRDRLIVAPSGA 60
 DB 1 MTKEEKPFGRLRSEFLPIHHEKLVLPMLMFCITENTYVLRDRLIVAPSGA 60
 QY 61 EAIPIKFWLVPPCAIIFMLIYAKLSNLSKQALFYAVGTPFLIFPFIYPLRDVL 120
 DB 61 EAIPIKFWLVPPCAIIFMLIYAKLSNLSKQALFYAVGTPFLIFPFIYPLRDVL 120
 QY 121 HPTFADRLQAILPGLIGVAILRNMTFAAFYVLAEIWSGWSLTMFWGPNETIKHE 180
 DB 121 HPTFADRLQAILPGLIGVAILRNMTFAAFYVLAEIWSGWSLTMFWGPNETIKHE 180
 QY 181 AKRYVALFGIGANISLASGAIYVASKLRSVSEGVDPWGISRLMAMTIVGLVMA 240
 DB 181 AKRYVALFGIGANISLASGAIYVASKLRSVSEGVDPWGISRLMAMTIVGLVMA 240
 QY 241 SYWINKNVLTDPFYPNPEEKQKGAAPKMNKDSFLYDRSPYILLTLVIAYGIC 300
 DB 241 SYWINKNVLTDPFYPNPEEKQKGAAPKMNKDSFLYDRSPYILLTLVIAYGIC 300
 QY 301 INLEVTWKSQKLQYPMNDYSEFGNFSFWTGVSVLIMLVGNAVIRKFGWLTGALV 360
 DB 301 INLEVTWKSQKLQYPMNDYSEFGNFSFWTGVSVLIMLVGNAVIRKFGWLTGALV 360
 QY 361 TPVVVLLTGIVFPLVIFRNQASGLVAMPFTPLMLLVVGAIONISKSKYALPSTK 420
 DB 361 TPVVVLLTGIVFPLVIFRNQASGLVAMPFTPLMLLVVGAIONISKSKYALPSTK 420
 QY 421 EMAYIPLDQEQKVGKAIIVVAPFKSGGALIQGILVYCGSIGAMTPYLAIVLFFII 480
 DB 421 EMAYIPLDQEQKVGKAIIVVAPFKSGGALIQGILVYCGSIGAMTPYLAIVLFFII 480
 QY 481 AIWIVSATKINKLFLAQSALKEQVADSDAPASS 515
 DB 481 AIWIVSATKINKLFLAQSALKEQVADSDAPASS 515
 RESULT 2
 TLCL CHLPRN STANDARD; PRT; 528 AA.
 AC Q84068;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE ADP carrier protein 1 (ADP/ATP translocase 1).
 GN TLCA OR CT065.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.
 NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UM-3/Ck;
 RX MEDLINE=9900809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AB001281; AAC67656.1; -

DR InterPro: IPR004667; ADP_ATP_car.

DR Pfam; PF03219; TLC; 1.

DR TIGRPFAM; TIGR00769; AAA; 1.

KW Transmembrane; Transport; ATP-binding; Multigene family;

KM Complete proteome.

FT TRANSMEM 24 44 POTENTIAL.

FT TRANSMEM 63 83 POTENTIAL.

FT TRANSMEM 93 113 POTENTIAL.

FT TRANSMEM 124 144 POTENTIAL.

FT TRANSMEM 149 169 POTENTIAL.

FT TRANSMEM 184 204 POTENTIAL.

FT TRANSMEM 220 240 POTENTIAL.

FT TRANSMEM 284 304 POTENTIAL.

FT TRANSMEM 327 347 POTENTIAL.

FT TRANSMEM 356 376 POTENTIAL.

FT TRANSMEM 381 401 POTENTIAL.

FT TRANSMEM 463 483 POTENTIAL.

SO SEQUENCE 528 AA; 58117 MW; EAD16A43B5B94B2 CRC64;

Query Match 6.4%; Score 33; DB 1; Length 528;
Best Local Similarity 100.0%; Pred. No. 1.6e-24;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 419 TKEMAYIPLDQEKVKGAIDVVAARFGKSG 451
Db 417 TKEMAYIPLDQEKVKGAIDVVAARFGKSG 449

RESULT 3
TLCL CHLMU STANDARD; PRT; 529 AA.

AC 09PKT5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE ADP ATP carrier protein 1 (ADP/ATP translocase 1).
GN TLCA OR TC0335.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_Taxid=83560;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / Ni99;
RX MEDLINE=20150255; PubMed=10684935; G11 S.R., Heidelberg J.F.,
RA Read T.D., Brumham R.C., Shen C., Utterback T., Berry K., Baas S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gaim M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AB002301; AAF39198.1; -
DR PIR; C81714; C81714.
DR TIGR; TC0335; -

DR InterPro: IPR004667; ADP_ATP_car.

DR Pfam; PF03219; TLC; 1.

DR TIGRPFAM; TIGR00769; AAA; 1.

KW Transmembrane; Transport; ATP-binding; Multigene family;

KM Complete proteome.

FT TRANSMEM 24 44 POTENTIAL.

FT TRANSMEM 63 83 POTENTIAL.

FT TRANSMEM 93 113 POTENTIAL.

FT TRANSMEM 124 144 POTENTIAL.

FT TRANSMEM 149 169 POTENTIAL.

FT TRANSMEM 184 204 POTENTIAL.

FT TRANSMEM 220 240 POTENTIAL.

FT TRANSMEM 284 304 POTENTIAL.

FT TRANSMEM 322 342 POTENTIAL.

FT TRANSMEM 356 376 POTENTIAL.

FT TRANSMEM 381 401 POTENTIAL.

FT TRANSMEM 463 483 POTENTIAL.

SO SEQUENCE 529 AA; 58304 MW; EED13A4C751071C9 CRC64;

Query Match 6.4%; Score 33; DB 1; Length 529;
Best Local Similarity 100.0%; Pred. No. 1.6e-24;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 419 TKEMAYIPLDQEKVKGAIDVVAARFGKSG 451
Db 417 TKEMAYIPLDQEKVKGAIDVVAARFGKSG 449

RESULT 4
TLCC RICPR STANDARD; PRT; 501 AA.

AC 09ZD5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ADP ATP carrier protein 3 (ADP/ATP translocase 3).
GN TLCC OR TLG3 OR Rp477.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; Rickettsiaceae; Rickettsia.
OX NCBI_Taxid=782;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893; Andersson J.O.,
RA Andersson S.G.E., Zomorodipour A., Almalk U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of mitochondria.";
RL Nature 396:133-140(1998).
CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.
CC THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF RICKETTSIAL PARASITISM.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; A135271; CAAL4932.1; -
DR PIR; B71707; B71707.
DR InterPro: IPR004667; ADP_ATP_car.
DR Pfam; PF03219; TLC; 1.
DR TIGRPFAM; TIGR00769; AAA; 1.
KW Transmembrane; Transport; ATP-binding; Multigene family;
KM Complete proteome.

SQ SEQUENCE 501 AA; 57182 MW; D1EC540E6D7E91F CRC64;
 Query Match 2.5%; Score 13; DB 1; Length 501;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 415 LFDSTKEMAYRPL 427
 DB 415 LFDSTKEMAYRPL 427
 RESULT 5
 TLIC2_ARATH STANDARD; PRT; 618 AA.
 AC P92935; Q9M933;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Chloroplast ADP/ATP carrier protein 2, chloroplast precursor (ADP/ATP translocase 2) (Adenine nucleotide translocase 2).
 GN ATP2 OR AT1G15500 OR T16N11.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RX MEDLINE=98206726; PubMed=11130712;
 RA Moehlmann T., Tjaden J., Schweoppe C., Winkler H.H., Kampfenkel K., Neuhaus H.E.;
 RT "Occurrence of two plastidic ATP/ADP transporters in Arabidopsis thaliana L. -- molecular characterization and comparative structural analysis of similar ATP/ADP translocators from plastids and Rickettsia prowazekii.";
 RT Eur. J. Biochem. 252:353-359 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Etnu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y., Gall J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huitzer L., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A., Miltchev J., Miranda M., Nguyen M., Niernan W.C., Osborne B.T., Pat G., Peterson J., Pham P.K., Rizzo W., Rooney T., Rowley D., Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M., Sun H., Taiton L.J., Tambunga G., Toriumi M.J., Town C.D., Utechtack T., Van Aken S., Vaysberg M., Vostokala V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RA "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";
 RT Nature 408:816-820 (2000).
 RL [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the SGP consortium (Salk/Stanford/PGEC).";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R., Feldmann K.;
 RT "Full-length cDNA from Arabidopsis thaliana.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST. INTEGRAL MEMBRANE PROTEIN.

CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLIC FAMILY.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts and other sequence problems.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X94626; CA64329.1; ALT SEQ.
 DR EMBL: AC013453; AAF71976.1; -
 DR EMBL: AY081350; AAL91239.1; -
 DR EMBL: AY128844; AAM91244.1; -
 DR EMBL: AY084374; AAM60955.1; -
 DR PIR: G86288; G86288.
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam: PF03219; TLIC_1.
 DR TIGRFAMs: TIGR00769; AAA_1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 FT TRANSIT 1
 FT CHAIN 1
 FT CHAIN 1
 SQ SEQUENCE 618 AA; 67530 MW; B2D49E77F17A867A CRC64;
 Query Match 2.1%; Score 11; DB 1; Length 618;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 432 KVGKKAIDV 442
 DB 509 KVGKKAIDV 519
 RESULT 6
 TLIC2_ARATH STANDARD; PRT; 624 AA.
 AC Q39002; Q94AN9; Q9C974;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Chloroplast ADP/ATP carrier protein 1, chloroplast precursor (ADP/ATP translocase 1) (Adenine nucleotide translocase 1).
 GN ATP1 OR AT1G80300 OR F516.5.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RX MEDLINE=96069943; PubMed=7589569;
 RA Kampfenkel K., Moehlmann T., Batz O., van Montagu M., Inze D., Neuhaus H.E.;
 RT "Molecular characterization of an Arabidopsis thaliana cDNA encoding a novel putative adenylate translocator of higher plants.";
 RT FEBS Lett. 374:351-355 (1995).
 RL [2]
 RP REVISIONS.
 RA Kampfenkel K.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Etnu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y., Gall J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huitzer L.,

RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hoyer S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Malti R., Matzili A.,
 RA Maltseva J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,
 RA Pal G., Peterson J., Plam P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Frazer C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana.";
 RT Nature 408:816-820(2000).
 RA [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Salk/Stanford/PGEC).";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP FUNCTION: AND SUBCELLULAR LOCATION.
 RX MEDLINE=9717797; PubMed=9025303;
 RA Neuhaus H.E., Thom E., Mohlmann T., Steup M., Kampfenkel K.;
 RT "Characterization of a novel eukaryotic ATP/ADP translocator located
 RT in the plastid envelope of Arabidopsis thaliana L.";
 RL Plant J. 11:73-82(1997).
 CC -1- FUNCTION: May function as an ATP importer.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST. INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z49227; CAAB9201.2; -;
 DR EMBL; AC018848; AAG52434.1; -;
 DR EMBL; AF428316; AAL16246.1; -;
 DR EMBL; AY045903; AAK76577.1; -;
 DR PIR; E96834; E96834.
 DR InterPro; IPR004667; ADP_ATP-car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 DR Transmembrane; Transport; ATP-binding; Multigene family;
 KM Chloroplast; Transit peptide
 FT TRANSIT 1 ?
 FT CHAIN ? 624
 FT TRANSMEM 108 128 CHLOROPLAST ADP/ATP CARRIER PROTEIN 1.
 FT TRANSMEM 182 202 POTENTIAL.
 FT TRANSMEM 240 260 POTENTIAL.
 FT TRANSMEM 315 335 POTENTIAL.
 FT TRANSMEM 446 466 POTENTIAL.
 FT TRANSMEM 545 565 POTENTIAL.
 FT CONFLICT 45 94
 FT CONFLICT 103 103
 FT CONFLICT 197 197
 FT CONFLICT 281 281
 SQ SEQUENCE 624 AA; 68134 MW; 530302A89AD08745 CRC64;
 Query Match 2.1%; Score 11; DB 1; Length 624;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 432 KVGKKAIDV 442
 DB 512 KVGKKAIDV 522

RESULT 7
 TLCD_SOLITU STANDARD; PRT; 631 AA.
 ID TLCD_SOLITU
 AC 024361;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Plastidic ATP/ADP-transporter.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4113;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Desiree;
 RA Neuhaus B.;
 RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RL -1- SUBCELLULAR LOCATION: CHLOROPLAST. INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y10821; CAA71785.1; -;
 DR PIR; T07420; T07420.
 DR InterPro; IPR004667; ADP_ATP-car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KM Transmembrane; Transport; ATP-binding; Multigene family;
 KM Chloroplast.
 SQ SEQUENCE 631 AA; 68886 MW; 33158380DE29D1EA CRC64;
 Query Match 2.1%; Score 11; DB 1; Length 631;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 432 KVGKKAIDV 442
 DB 510 KVGKKAIDV 520

RESULT 8
 TLCD_RICPR STANDARD; PRT; 512 AA.
 ID TLCD_RICPR
 AC Q92D47;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP/ATP carrier protein 4 (ADP/ATP translocase 4).
 GN TLCD OR TLC4 OR RP500.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sichelberger-Ponten T., Almark U.C.M., Podowski R.M., Naeelund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurliand C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RT Nature 396:133-140(1998).
 CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
 CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.

THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF RICKETTSIAL PARASITISM.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLUCASE TLC FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AJ235272; CAA14952.1; -

DR PIR: F71653; F71653; ADP_ATP_car.

DR InterPro: IPR004667; ADP_ATP_car.

DR Pfam: PF03219; TLC; 1.

DR TIGRFAMs: TIGR00769; AAA; 1.

KM Transmembrane; Transport; ATP-binding; Multigene family;

KM Complete proteome.

SO SEQUENCE 512 AA; 56208 MW; D09A265DC38936E8 CRC64;

Query Match 1.7%; Score 9; DB 1; Length 498;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

415 LFDSTKEMAY 424
|||||
422 LFDSTKEMAY 431

RESULT 9

TLC_RICPR STANDARD; PRT; 498 AA.

AC P19568;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ADP carrier protein 1 (ADP/ATP translocase 1).
GN TLC OR TLC1 OR TLC OR RP053.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX NCBI_Taxid=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=90060776; PubMed=2555259;
RA Williamson L.R., Plano G.V., Winkler H.H., Krause D.C., Wood D.O.;
RT "Nucleotide sequence of the Rickettsia prowazekii ATP/ADP translocase-
RT encoding gene." (1989).
RL Gene 80:269-278(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
RA Schenker-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria." (1998).
RL Nature 396:133-140(1998).
RN [3]
RP TOPOLOGY
RX MEDLINE=91258319; PubMed=1904433;
RA Plano G.V., Winkler H.H.;
RT "Identification and initial topological analysis of the Rickettsia
RT prowazekii ATP/ADP translocase." (1991).
RL J. Bacteriol. 173:3389-3396(1991).
RN [4]
RP TOPOLOGY
RX MEDLINE=90334349; PubMed=2165754;
RA Plano G.V., Wood D.O., Winkler H.H.;

"Rickettsia prowazekii and ATP/ADP translocase. Analysis of gene
functions encoding beta-galactosidase-ATP/ADP translocase fusion
proteins." (1990).
RL Ann. N.Y. Acad. Sci. 590:397-407(1990).
RN [5]
RP TOPOLOGY
RX MEDLINE=99141397; PubMed=9917392;
RA Alexeyev M.F., Winkler H.H.;
RT "Membrane topology of the Rickettsia prowazekii ATP/ADP translocase
RT revealed by novel dual pho-lac reporters." (1999).
RL J. Mol. Biol. 285:1503-1513(1999).
CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.
CC THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF
CC RICKETTSIAL PARASITISM.
CC RICKETTSIAL PARASITISM.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLUCASE TLC FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M28816; AAA26382.1; -

DR EMBL: AJ235270; CAA14524.1; -

DR PIR: JQ0026; JQ0026.

DR InterPro: IPR004667; ADP_ATP_car.

DR Pfam: PF03219; TLC; 1.

DR TIGRFAMs: TIGR00769; AAA; 1.

KM Transmembrane; Transport; ATP-binding; Multigene family;

KM Complete proteome.

FT DOMAIN 1 33 CYTOPLASMIC (PROBABLE).
FT TRANSSEM 34 54 PROBABLE.
FT DOMAIN 55 67 EXTRACELLULAR (PROBABLE).
FT TRANSSEM 68 88 PROBABLE.
FT DOMAIN 89 92 CYTOPLASMIC (PROBABLE).
FT TRANSSEM 93 113 PROBABLE.
FT DOMAIN 114 147 EXTRACELLULAR (PROBABLE).
FT TRANSSEM 148 168 PROBABLE.
FT DOMAIN 169 184 CYTOPLASMIC (PROBABLE).
FT TRANSSEM 185 205 PROBABLE.
FT DOMAIN 206 218 EXTRACELLULAR (PROBABLE).
FT TRANSSEM 219 239 PROBABLE.
FT DOMAIN 240 279 CYTOPLASMIC (PROBABLE).
FT TRANSSEM 280 300 PROBABLE.
FT DOMAIN 301 320 EXTRACELLULAR (PROBABLE).
FT TRANSSEM 321 341 PROBABLE.
FT DOMAIN 342 348 CYTOPLASMIC (PROBABLE).
FT TRANSSEM 349 369 PROBABLE.
FT DOMAIN 370 379 EXTRACELLULAR (PROBABLE).
FT TRANSSEM 380 400 PROBABLE.
FT DOMAIN 401 438 CYTOPLASMIC (PROBABLE).
FT TRANSSEM 439 459 PROBABLE.
FT DOMAIN 460 465 EXTRACELLULAR (PROBABLE).
FT TRANSSEM 466 486 PROBABLE.
FT DOMAIN 487 498 CYTOPLASMIC (PROBABLE).
FT DISULFID 37 85 PROBABLE.
FT NP BIND 436 442 ATP (POTENTIAL).
SO SEQUENCE 498 AA; 56821 MW; D3C450D2BC0BE53E CRC64;

Query Match 1.7%; Score 9; DB 1; Length 498;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

246 NKNVLTDP 254
|||||
241 NKNVLTDP 249

RESULT 10

TCB2_MOUSE
ID TCB2_MOUSE STANDARD; PRT; 173 AA.
AC P01851;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 21-JUL-1986 (Rel. 01, Last annotation update)
DE T-cell receptor beta-2 chain C region.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B10.A;
RX MEDLINE=84270704; PubMed=6336329;
RA Gascogne N.R.J., Chien Y., Becker D.M., Kavalier J., Davis M.M.;
RT "Genomic organization and sequence of T-cell receptor beta-chain
constant- and joining-region genes."
RL Nature 310:387-391(1984).
RN [2]
RP SEQUENCE FROM N.A. (CLONE 2C).
RC STRAIN=BALB.B;
RX MEDLINE=84245824; PubMed=6330561;
RA Saito H., Kranz D.M., Takagaki Y., Hayday A.C., Eisen H.N.,
RA Tonegawa S.;
RT "Complete primary structure of a heterodimeric T-cell receptor
deduced from cDNA sequences."
RL Nature 309:757-762(1984).
CC -1- MISCELLANEOUS: CLONE B10.A WAS ISOLATED FROM A CYTOTOXIC T
CC LYMPHOCYTE.
CC -1- MISCELLANEOUS: CLONE 2C WAS ISOLATED FROM A CYTOTOXIC T
CC LYMPHOCYTE.
CC HSPSP; P01842; 2MCG.
DR InterPro; IPR007110; Ig_1-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; Igcl; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW T-cell; Receptor; Immunoglobulin domain; Transmembrane; Glycoprotein.
FT T-CELL 1
FT NON_TER 1
FT DOMAIN 1 146 C REGION.
FT TRANSMEM 147 168
FT DOMAIN 169 173 CYTOPLASMIC TAIL.
FT VARIANT 50 50 K->R (IN CLONE 2C).
FT VARIANT 70 70 Y->H (IN CLONE 2C).
SQ SEQUENCE 173 AA; 19297 MW; A5458149614CF295 CRC64;
Query Match 1.6%; Score 8; DB 1; Length 173;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 233 VSGGLVLA 240
DB 159 VSGGLVLA 166
RESULT 11
O2H2_HUMAN STANDARD; PRT; 312 AA.
ID O2H2_HUMAN
AC O95918;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Olfactory receptor 2H2 (Hs6M1-12).
GN OR2H2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Younger R.;

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL031983; CAA21455.1; -
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHOPOPSN.
DR PROSITE; PS00237; G_PROTEIN_REC_P1_1; 1.
DR PROSITE; PS0262; G_PROTEIN_REC_P1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction.
FT DOMAIN 1 23
FT TRANSMEM 24 47
FT DOMAIN 48 55
FT TRANSMEM 56 77
FT DOMAIN 78 98
FT TRANSMEM 99 118
FT DOMAIN 119 137
FT TRANSMEM 138 156
FT DOMAIN 157 193
FT TRANSMEM 194 217
FT DOMAIN 218 234
FT TRANSMEM 235 257
FT DOMAIN 258 270
FT TRANSMEM 271 290
FT DOMAIN 291 312
FT DISULFID 95 187
FT CARBOHYD 3 3
SQ SEQUENCE 312 AA; 34797 MW; 2C7AB6C7E08AB8B CRC64;
Query Match 1.6%; Score 8; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 94 LFYAVGTP 101
DB 274 LFYAVGTP 281
RESULT 12
O2H1_HUMAN STANDARD; PRT; 316 AA.
ID O2H1_HUMAN
AC O9GZK4; O43629; O43661; O9GZK9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Olfactory receptor 2H1 (Hs6M1-16) (Olfactory receptor 6-2) (ORE-2)
GN OLFRA2A-9004.14/9026.2).
GN OR2H1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND VARIANT ASN-63.
RA Ziegler A., Ehlers A., Forbes S.A., Trowsdale J., Uchanska-Ziegler B.,
RA Voiz A., Younger R., Beck S.;
RT "Polymorphic olfactory receptor genes and HLA loci constitute extended
RT haplotypes."
RL (in) Kasahara M. (eds.);
RL Major histocompatibility complex-evolution, structure, and function,
RL pp.110-130, Springer-Verlag, Tokyo (2000).
RN [2]

RP SEQUENCE FROM N.A.
RC Tissue-Testis:
RA Volz A., Ehlers A., Younger R., Forbes S., Trowdale J., Beck S.,
Ziegler A.;
RT "Complex transcriptional control of MHC-linked olfactory receptor
genes includes long distance and extensive alternative splicing, exon
RT shunting and premature polyadenylation.";
RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Whitaker H.;
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 58-251 FROM N.A.
RA Galliano H.;
RT "Olfactory receptor gene cluster in man and mouse major
RT histocompatibility complex.";
RL Immunogenetics 0:0-0(1998).
RN [5]
RP SEQUENCE OF 58-251 FROM N.A.
RA Amadou C., Avouet P., Ribouchon M.-T., Bouissou C., Tazi-Ahmini R.,
Ayer C., Pontarotti P.;
RL Submitted (Feb-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ302604; CAC20524.1; -
DR EMBL; AJ302605; CAC20525.1; -
DR EMBL; AJ302606; CAC20526.1; -
DR EMBL; AJ302607; CAC20527.1; -
DR EMBL; AJ302608; CAC20528.1; -
DR EMBL; AJ302609; CAC20529.1; -
DR EMBL; AJ302610; CAC20530.1; -
DR EMBL; AJ302611; CAC20531.1; -
DR EMBL; AJ302612; CAC20532.1; -
DR EMBL; AJ302613; CAC20533.1; -
DR EMBL; AJ459847; CAD31039.1; -
DR EMBL; AJ459847; CAD31039.1; -
DR EMBL; AJ459847; CAD31039.1; -
DR EMBL; AF042078; AAC00184.1; -
DR EMBL; AF044491; AAC00184.1; -
DR GeneW; HGNC:8252; OR2H1.
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHOPOPSN.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL 1; 1.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL 2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction; Polymorphism.
FT VARIANT 63 63 D->N (IN ALLELE 6M1-16-02).
FT FT /FTID=VAR 010944.
FT CONFLICT 158 158 T->K (IN REF. 5).
FT CONFLICT 184 184 R->G (IN REF. 5).
FT CONFLICT 215 215 S->Y (IN REF. 5).
SQ SEQUENCE 316 AA; 35338 MW; 19140A7FD690CF85 CRC64;
Query Match 1.6%; Score 8; DB 1; Length 316;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 13
C0H3 HUMAN
ID 02H3 HUMAN STANDARD; PRT; 316 AA.
AC Q15062;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Olfactory receptor 2H3 (Olfactory receptor-like protein FAT11).
GN OR2H3 OR OLFR2 OR FAT11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95394447; PubMed=7665158;
RA Fan W., Liu Y.-C., Paximou S., Weissman S.M.;
RT "Olfactory receptor-like genes are located in the human major
RT histocompatibility complex.";
RL Genomics 27:119-123(1995).
RN [2]
RP SEQUENCE OF 10-312 FROM N.A., AND VARIANTS SER-30; VAL-48 AND ALA-220.
RX MEDLINE=20341390; PubMed=10880742;
RA Eklund A.C., Belchak M.M., Lapidos K., Raha-Chowdhury R., Ober C.;
RT "Polymorphisms in the HLA-linked olfactory receptor genes in the
RT Hutterites.";
RL Hum. Immunol. 61:711-717(2000).
CC -1- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L35475; AAB36567.1; -
DR EMBL; AF211939; AAF98751.1; -
DR EMBL; AF211940; AAF98752.1; -
DR EMBL; AF211941; AAF98753.1; -
DR EMBL; AF211942; AAF98754.1; -
DR PIR; A57069; A57069.
DR GeneW; HGNC:8254; OR2H3.
DR MIM; 600578; -
DR GO; GO:0003793; F:defense/immunity protein activity; TAS.
DR GO; GO:0004984; F:olfactory receptor activity; TAS.
DR GO; GO:0007618; P:maturing; TAS.
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHOPOPSN.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL 1; 1.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL 2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction.
FT DOMAIN 1 22 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 23 46 1 (POTENTIAL).
FT DOMAIN 47 55 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 56 77 2 (POTENTIAL).
FT DOMAIN 78 98 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 99 118 3 (POTENTIAL).
FT DOMAIN 119 137 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 138 158 4 (POTENTIAL).
FT DOMAIN 159 198 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 199 220 5 (POTENTIAL).
FT DOMAIN 221 234 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 235 259 6 (POTENTIAL).
FT DOMAIN 260 270 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 271 290 7 (POTENTIAL).
FT DOMAIN 291 316 CYTOPLASMIC (POTENTIAL).
FT DISULFID 95 187 BY SIMILARITY.

FT CARBOHYD 3 3 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARIANT 30 30 F->S.
 FT VARIANT 48 48 /FTID=VAR_010227.
 FT VARIANT 220 220 A->V.
 FT VARIANT 220 220 /FTID=VAR_010228.
 FT CONFLICT 22 22 E->G (IN REF. 1).
 FT CONFLICT 28 28 V->D (IN REF. 1).
 FT CONFLICT 30 30 F->I (IN REF. 1).
 FT CONFLICT 51 51 P->T (IN REF. 1).
 FT CONFLICT 81 81 V->A (IN REF. 1).
 FT CONFLICT 113 114 LT->MK (IN REF. 1).
 FT CONFLICT 153 153 E->G (IN REF. 1).
 FT CONFLICT 230 232 KGR->TAM (IN REF. 1).
 FT CONFLICT 267 267 E->G (IN REF. 1).
 FT CONFLICT 287 287 I->V (IN REF. 1).
 FT CONFLICT 295 296 VT->IK (IN REF. 1).
 FT CONFLICT 299 299 F->L (IN REF. 1).
 FT CONFLICT 307 311 RDSRE->MGLO (IN REF. 2).
 SQ SEQUENCE 316 AA; 35410 MM; BE29B672B663A7CB CRC64;
 Query Match 1.6%; Score 8; DB 1; Length 316;
 Best Local Similarity 100.0%; Pred. No. 6.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 94 LFYAVGTP 101
 DB 274 LFYAVGTP 281
 RESULT 14
 ACRA_ECOLI STANDARD; PRT; 400 AA.
 ID ACRA_ECOLI
 AC P1504; P78188; Q59386;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DN Acetate kinase (EC 2.7.2.1) (acetate kinase).
 OS ACRA OR ACK OR B2296 OR C2838 OR Z3558 OR ECS3180.
 OS Escherichia coli.
 OS Escherichia coli O6, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OC NCBI_TaxID=562, 217992, 83334;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=K12;
 RX MEDLINE=89123075; PubMed=2536666;
 RA Matsuyama A., Yamamoto H., Nakano E.;
 RT "Cloning, expression, and nucleotide sequence of the Escherichia coli
 RT K-12 ackA gene.";
 RL J. Bacteriol. 171:577-580(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / KH131;
 RX MEDLINE=95189796; PubMed=7883769;
 RA Kakuda H., Hosono K., Shirotani K., Ichihara S.;
 RT "Identification and characterization of the ackA (acetate kinase
 RT A)-pta (phosphotransacetylase) operon and complementation analysis of
 RT acetate utilization by an ackA-pta deletion mutant of Escherichia
 RT coli.";
 RL J. Biochem. 116:916-922(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97349980; PubMed=9205837;
 RA Yamamoto Y., Alba H., Baba T., Hayaishi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakase S., Nakamura Y., Nishimoto H.,
 RA Oshima T., Oyama S., Saito N., Sampei G., Satch Y., Sivasubram S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horiiuchi T.;
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
 RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:1H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga H.;
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC -1- FUNCTION: INVOLVED IN THE ACTIVATION OF ACETATE TO ACETYL COA AND
 CC THE SECRETION OF ACETATE. DURING ANAEROBIC GROWTH OF THE ORGANISM,
 CC THIS ENZYME IS ALSO INVOLVED IN THE SYNTHESIS OF MOST OF THE ATP
 CC FORMED CATABOLICALLY.
 CC -1- CATALYTIC ACTIVITY: ATP + acetate = ADP + acetyl phosphate.
 CC -1- PATHWAY: Conversion of acetate to acetyl-CoA, first step.
 CC -1- SUBUNIT: Homodimer.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to the acetate kinase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL: M22956; AAA23406.1; -;
 DR EMBL: D17576; BAA04501.1; -;
 DR EMBL: AB000318; AAC75356.1; -;
 DR EMBL: D90861; BAA16135.1; -;
 DR EMBL: D90860; BAA16131.1; -;
 DR EMBL: AB016763; AAN81292.1; -;

DR EMBL; AB005461; AAG57425.1; -
 DR EMBL; AP002561; BAB36603.1; -
 DR PIR; D91026; D91026.
 DR PIR; J04048; KIECAA.
 DR PDB; 1LRG; 19-JUN-02.
 DR SWISS-2DPAGE; P15046; COLI.
 DR ECODBASE; G041.3; 6TH EDITION.
 DR Ecogene; EGI0027; ackA.
 DR HAMAP; MF_00020; -; 1.
 DR InterPro; IPR000890; Acetate_kin.
 DR InterPro; IPR004372; AckA.
 DR Pfam; PF00871; Acetate_kinase; 1.
 DR PRINTS; PR00471; ACETATEKINASE.
 DR TIGRFAMs; TIGR00016; ackA; 1.
 DR PROSITE; PS01075; ACETATE_KINASE_1; 1.
 DR PROSITE; PS01076; ACETATE_KINASE_2; 1.
 DR TRANSFERASE; Kinase; Complete proteome; 3D-structure.
 DR CONFLICT 221 221 R -> G (IN REF. 2).
 DR CONFLICT 240 240 M -> I (IN REF. 2).
 FT SEQUENCE 400 AA; 43290 MW; 709C855C514656C8 CRC64;
 SO

Query Match 1.6%; Score 8; DB 1; Length 400;
 Best Local Similarity 100.0%; Pred. No. 8.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 443 AARFGKSG 450
 Db 363 AARFGKSG 370

RESULT 15
 ID ACKA_SALTY STANDARD; PRT; 400 AA.
 AC Q8XFG9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Acetate kinase (EC 2.7.2.1) (Acetokinase).
 GN ACKA OR STM2337 OR STY2567 OR T0527.
 OS Salmoneilla typhimurium, and
 OS Salmoneilla typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmoneilla.
 OX NCBI_TaxID=602, 601;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Portolillo S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmoneilla enterica serovar Typhimurium
 LT2";
 RT Nature 413:852-856(2001).
 RL [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. typhi; STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Comerton P.,
 RA Cronin A., Davis P., Davies R.W., Dowd L., White N., Farrar J.,
 RA Feltham T., Hamlin N., Haque A., Hen T.T., Holroyd S., Jagsels K.,
 RA Krogh A., Larsen T.S., Leach S., Moule S., O'Garra P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RA "Complete genome sequence of a multiple drug resistant Salmoneilla
 RT enterica serovar Typhi CT18";
 RL Nature 413:848-852(2001).
 RN [3]
 RP SEQUENCE FROM N.A.

RC SPECIES=S. typhi; STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Lion S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodyanski V., Schwartz D.C., Blattner P.R.;
 RT "Comparative genomics of Salmoneilla enterica serovar Typhi strains Ty2
 and CT18";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- CATALYTIC ACTIVITY: ATP + acetate = ADP + acetyl phosphate.
 CC -1- PATHWAY: Conversion of acetate to acetyl-coA, first step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the acetokinase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 or send an email to license@ebi.ac.uk).
 CC -----

DR EMBL; AB008805; AA21238.1; -
 DR EMBL; AL627274; CAD07569.1; -
 DR EMBL; AE016836; AA068233.1; -
 DR HAMAP; MF_00020; -; 1.
 DR InterPro; IPR000890; Acetate_kin.
 DR InterPro; IPR004372; AckA.
 DR Pfam; PF00871; Acetate_kinase; 1.
 DR PRINTS; PR00471; ACETATEKINASE.
 DR TIGRFAMs; TIGR00016; ackA; 1.
 DR PROSITE; PS01075; ACETATE_KINASE_1; 1.
 DR PROSITE; PS01076; ACETATE_KINASE_2; 1.
 DR TRANSFERASE; Kinase; Complete proteome.
 SO SEQUENCE 400 AA; 43257 MW; BD26BC1AB08CA7AD CRC64;

Query Match 1.6%; Score 8; DB 1; Length 400;
 Best Local Similarity 100.0%; Pred. No. 8.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 443 AARFGKSG 450
 Db 363 AARFGKSG 370

RESULT 16
 ID ACKA_YERPE STANDARD; PRT; 400 AA.
 AC Q8ZDU6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Acetate kinase (EC 2.7.2.1) (Acetokinase).
 GN ACKA OR YFO2566 OR Y1622.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Terraza A.M.,
 RA Chillingworth T., Cronin A., Davies R.W., Davis P., Dougan G.,
 RA Feltham T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
 RA Leach S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
 RA "Genome sequence of Yersinia pestis, the causative agent of plague";
 RT Nature 413:523-527(2001).
 RL [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIM5 / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;

RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhong S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,
 RA Perry R.D.;
 "Genome sequence of *Yersinia pestis* KIM";
 RL J. Bacteriol. 184:4601-4611 (2002).
 CC -1- CATALYTIC ACTIVITY: ATP + acetate = ADP + acetyl phosphate.
 CC -1- PATHWAY: Conversion of acetate to acetyl-CoA, first step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the acetate kinase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ414152; CAC91368.1; -;
 CC EMBL: AE013765; AAM85191.1; -;
 CC PIR: AD0313; AD0313.
 DR HAMAP: MF_00020; -; 1.
 DR InterPro: IPR0000890; Acetate_kin.
 DR InterPro: IPR004172; ACKA.
 DR Pfam: PF00871; Acetate_Kinase; 1.
 DR PRINTS: PR00471; ACETATE_KINASE.
 DR TIGRFAMs: TIGR00016; ACKA; 1.
 DR PROSITE: PS01075; ACETATE_KINASE_1; 1.
 DR PROSITE: PS01076; ACETATE_KINASE_2; 1.
 KW Transferase; Kinase; Complete proteome.
 SQ SEQUENCE 400 AA; 43018 MW; 21BA6CC5B1A27716 CRC64;
 Query Match 1.6%; Score 8; DB 1; Length 400;
 Best Local Similarity 100.0%; Pred. No. 8.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 443 AARFGKSG 450
 DB 363 AARFGKSG 370
 RESULT 17
 CRTI_SRYN3
 ID CRTI_SRYN3 STANDARD; PRT; 472 AA.
 AC P29273;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
 GN PDS OR CRD OR SLR1254.
 OS *Synechocystis* sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; *Synechocystis*.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92256820; PubMed=1581575;
 RA Martinez-Ferez I.M., Vioque A.;
 RT "Nucleotide sequence of the phytoene desaturase gene from
 RT *Synechocystis* sp. PCC 6803 and characterization of a new mutation
 RT which confers resistance to the herbicide norflurazon.";
 RL Plant Mol. Biol. 18:981-983 (1992).
 RN [2]
 RP REVISIONS TO C-TERMINUS.
 RA Vioque A.;
 RL Submitted (NOV-1993) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K.,

RA Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A.,
 RA Yamada M., Yasuda M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions";
 RL DNA Res. 3:109-136 (1996).
 CC -1- FUNCTION: THIS ENZYME CONVERTS PHYTOENE INTO ZETA-CAROTENE VIA THE
 CC INTERMEDIARY OF PHYTOFLUENE BY THE SYMMETRICAL INTRODUCTION OF TWO
 CC DOUBLE BONDS AT THE C-11 AND C-11' POSITIONS OF PHYTOENE.
 CC -1- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
 CC -1- ENZYME REGULATION: INHIBITED BY THE HERBICIDE NORFLURAZON IN A
 CC NON-COMPETITIVE WAY.
 CC -1- PATHWAY: Carotenoid biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Membrane-associated (Probable).
 CC -1- SIMILARITY: BELONGS TO THE PHYTOENE DEHYDROGENASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X62574; CAA44452.1; -;
 CC EMBL: D90809; BAA17847.1; -;
 CC PIR: S74886; S74886.
 DR InterPro: IPR000759; Adrxn_reductase.
 DR InterPro: IPR001613; Amineoxid fl.
 DR InterPro: IPR002937; Amine oxidase.
 DR Pfam: PF01593; Amine oxidase; 1.
 DR PRINTS: PR00419; ADXRDPASE.
 DR PRINTS: PR00757; AMINEOXIDASE.
 KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;
 KW Membrane; Herbicide resistance; Complete proteome.
 FT NP BIND 7 23 FAD (ADP PART) (POTENTIAL).
 FT VARIANT 195 195 R -> C (CONFERS RESISTANCE TO THE
 FT SEQUENCE 472 AA; 52920 MW; 81D089A6DMA28758 CRC64;
 Query Match 1.6%; Score 8; DB 1; Length 472;
 Best Local Similarity 100.0%; Pred. No. 9.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 139 GLVAILRN 146
 DB 109 GLVAILRN 116
 RESULT 18
 ANGT_HUMAN
 ID ANGT_HUMAN STANDARD; PRT; 485 AA.
 AC P01019; Q16358; Q16359; Q96F91;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Angiotensinogen precursor [Contains: Angiotensin I (Ang I);
 DE Angiotensin II (Ang II); Angiotensin III (Ang III) (Des-Asp1)] -
 DE Angiotensin II].
 GN AGT OR SERPINB8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89170129; PubMed=2924688;
 RA Galliard I., Clauser E., Corvol P.;
 RT "Structure of human angiotensinogen gene.";
 RL DNA 8:87-99 (1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85000455; PubMed=6089875;

RA Kageyama R., Ohkubo H., Nakanishi S.;
 RT "Primary structure of human preangiotensinogen deduced from the
 RT cloned cDNA sequence.";
 RL Biochemistry 23:3603-3609 (1984).
 RN [3]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=90237063; PubMed=1692023;
 RA Fukumizu A., Takahashi S., Seo M.S., Tada M., Tanimoto K., Uehara S.,
 RA Murakami K.;
 RT "Structure and expression of the human angiotensinogen gene.
 RT Identification of a unique and highly active promoter.";
 RL J. Biol. Chem. 265:7576-7582 (1990).
 RN [4]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=2338257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueidl T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Murzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buterfield V.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [5]
 RN SEQUENCE OF 1-338 FROM N.A.
 RX MEDLINE=87244745; PubMed=2885106;
 RA Kunapuli S.P., Kumar A.;
 RT "Molecular cloning of human angiotensinogen cDNA and evidence for the
 RT presence of its mRNA in rat heart.";
 RL Circ. Res. 60:786-790 (1987).
 RN [6]
 RN SEQUENCE OF 34-45, AND SUBUNITS.
 RC TISSUE=Serum;
 RX MEDLINE=95293954; PubMed=7539791;
 RA Oxyg C., Haaning J., Kristensen L., Wagner J.M., Rubin I.,
 RA Stigbrand T., Gleich G.J., Sottrup-Jensen L.;
 RT "Identification of angiotensinogen and complement C3dg as novel
 RT proteins binding the proform of eosinophil major basic protein in
 RT human pregnancy serum and plasma.";
 RL J. Biol. Chem. 270:13645-13651 (1995).
 RN [7]
 RN SEQUENCE OF 34-43.
 RX MEDLINE=69014170; PubMed=4300938;
 RA Arkawa K., Minohara A., Yamada J., Nakamura M.;
 RT "Enzymatic degradation and electrophoresis of human angiotensin I.";
 RL Biochim. Biophys. Acta 168:106-112 (1968).
 RN [8]
 RN CARBOHYDRATE-LINKAGE SITES.
 RX MEDLINE=86056581; PubMed=3934016;
 RA Campbell D.J., Bounhik J., Coezy E., Menard J., Corvol P.;
 RT "Processing of rat and human angiotensinogen precursors by microsomal
 RT membranes.";
 RL Mol. Cell. Endocrinol. 43:31-40 (1985).
 RN [9]
 RN FUNCTION OF ANGIOTENSIN III.
 RX MEDLINE=7516649; PubMed=1132082;
 RA Goodfriend T.L., Peach M.J.;
 RT "Angiotensin III: (DBS-Aspartic Acid-1)-Angiotensin II. Evidence and
 RT speculation for its role as an important agonist in the renin -
 RT angiotensin system.";

RL Circ. Res. 36:38-48 (1975).
 RN [10]
 RN STRUCTURE BY NMR OF ANGIOTENSIN II.
 RX MEDLINE=98151281; PubMed=9492317;
 RA Carpenter K.A., Wilkes B.C., Schiller P.W.;
 RT "The octapeptide angiotensin II adopts a well-defined structure in a
 RT phospholipid environment.";
 RL Eur. J. Biochem. 251:448-453 (1998).
 RN [11]
 RN VARIANTS MET-207, THR-268 AND CYS-281.
 RX MEDLINE=93008239; PubMed=1394429;
 RA Jeunemaitre X., Soubrier F., Kotelevtsev Y.V., Lifton R.P.,
 RA Williams C.S., Charu A., Hunt S.C., Hopkins P.N., Williams R.R.,
 RA Lalouel J.-M., Corvol P.;
 RT "Molecular basis of human hypertension: role of angiotensinogen.";
 RL Cell 71:169-180 (1992).
 RN [12]
 RN VARIANT THR-268.
 RX MEDLINE=93291876; PubMed=8513325;
 RA Ward K., Hata A., Jeunemaitre X., Helin C., Nelson L., Nankawa C.,
 RA Farrington P.F., Ogasawara M., Suzumori K., Tomoda S., Berzbi S.,
 RA Saeki M., Corvol P., Lifton R.P., Lalouel J.-M.;
 RT "A molecular variant of angiotensinogen associated with
 RT preclampsia.";
 RL Nat. Genet. 4:59-61 (1993).
 RN [13]
 RN VARIANTS ILE-242; ARG-244 AND CYS-281.
 RX MEDLINE=95331754; PubMed=7607642;
 RA Hixson J.E., Powers P.K.;
 RT "Detection and characterization of new mutations in the human
 RT angiotensinogen gene (AGT).";
 RL Hum. Genet. 96:110-112 (1995).
 RN [14]
 RN CHARACTERIZATION OF VARIANT CYS-281.
 RX MEDLINE=96192253; PubMed=8621667;
 RA Gimenez-Roqueplo A.P., Lecoq I., Cohen P., Simon D., Guyene T.T.,
 RA Celierier J., Pau B., Corvol P., Clausen E., Jeunemaitre X.;
 RT "The natural mutation Y242C of human angiotensinogen leads to abnormal
 RT glycosylation and altered immunological recognition of the protein.";
 RL J. Biol. Chem. 271:9838-9844 (1996).
 RN [15]
 RN -1- FUNCTION: IN RESPONSE TO LOWERED BLOOD PRESSURE, THE ENZYME RENIN
 CC CLEAVES ANGIOTENSIN I, FROM ANGIOTENSINOGEN. ACE (ANGIOTENSIN
 CC CONVERTING ENZYME) THEN REMOVES A DIPEPTIDE TO YIELD THE
 CC PHYSIOLOGICALLY ACTIVE PEPTIDE ANGIOTENSIN II, THE MOST POTENT
 CC PRESSOR SUBSTANCE KNOWN, WHICH HELPS REGULATE VOLUME AND MINERAL
 CC BALANCE OF BODY FLUIDS.
 CC -1- FUNCTION: Angiotensin III stimulates aldosterone release.
 CC -1- SUBUNIT: During pregnancy, exists as a disulfide-linked 2:2
 CC heterotrimer with the proform of PRG2 and as a complex (probably
 CC a 2:2:2 heterotrimer) with pro-PRG2 and C3dg.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Synthesized by the liver and secreted in the
 CC plasma.
 CC -1- DISEASE: AGT SEEMS TO BE ASSOCIATED WITH A PREDISPOSITION TO
 CC ESSENTIAL HYPERTENSION AS WELL AS PREGNANCY-INDUCED HYPERTENSION
 CC (PIH) (PRECLAMPSIA).
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-10 IS THE INITIATOR.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: K02215; AAA51731.1; -;
 CC EMBL: M24689; AAA51679.1; -;
 CC EMBL: M24686; AAA51679.1; JOINED.
 CC EMBL: M24687; AAA51679.1; JOINED.
 CC EMBL: M24688; AAA51679.1; JOINED.
 CC EMBL: X15324; CAA33385.1; -;

DR EMBL: X15325; CAA33385.1; JOINED.
 DR EMBL: X15326; CAA33385.1; JOINED.
 DR EMBL: X15327; CAA33385.1; JOINED.
 DR EMBL: M69110; AAA52282.1; -.
 DR EMBL: BC011519; AAH11519.1; -.
 DR EMBL: S78529; AAD14287.1; -.
 DR EMBL: S78530; AAD14288.1; -.
 DR PIR: A35203; ANHU.
 DR SWISS-2DPAGE; P01019; HUMAN.
 DR Genew; HGNC:333; AGT.
 DR MIM; 106150; -.
 DR GO; GO:0005625; C:soluble fraction; TAS.
 DR GO; GO:0004667; F:serine protease inhibitor activity; TAS.
 DR GO; GO:0007166; P:cell surface receptor linked signal transdu. . .; TAS.
 DR GO; GO:0007677; P:cell-cell signalling; TAS.
 DR GO; GO:0007565; P:pregnancy; TAS.
 DR GO; GO:0008217; P:regulation of blood pressure; TAS.
 DR InterPro; IPR000227; Angiotensin.
 DR Pfam; P000079; Serpin; 1.
 DR PRINTS; P000654; ANGIOTENSIN.
 DR SMART; S000093; SERPIN; 1.
 DR PROSITE; PS00284; SERPIN; 1.
 DR KMW; Vasoconstrictor; Glycoprotein; Plasma; Serpin; Signal;
 DR Disease mutation; Polymorphism.
 FT SIGNAL 1 33
 FT CHAIN 1 485
 FT PEPTIDE 34 43 ANGIOTENSIN I.
 FT PEPTIDE 34 41 ANGIOTENSIN II.
 FT PEPTIDE 35 41 ANGIOTENSIN III.
 FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .).
 Query Match 1.6%; Score 8; DB 1; Length 485;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 ADRLQAIL 133
 Db 155 ADRLQAIL 162

RESULT 19
 NTC4_MOUSE
 ID NTC4_MOUSE STANDARD; PRT; 1964 AA.
 AC P31695; O35442; O88314; O88316; Q62389; Q62390; Q9R1W9; Q9R1X0;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neogenic locus notch homolog protein 4 precursor (Notch 4)
 GN [contains: Transforming protein Int-3].
 OS NOTCH4 OR INT3 OR INT-3.
 OC Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92194507; PubMed=1312643;
 RA Robbins J., Blondel B.J., Gallahan D., Callahan R.;
 RT "mouse mammary tumor gene int-3: a member of the notch gene family
 RT transforms mammary epithelial cells."; J. Virol. 66:2594-2599(1992).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RX MEDLINE=97294559; PubMed=9150355;
 RA Callahan D., Callahan R.;
 RT "The mouse mammary tumor associated gene INT3 is a unique member of
 RT the NOTCH gene family (NOTCH4)."; Oncogene 14:1883-1890(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung, and Testis;

RX MEDLINE=96281668; PubMed=8681805;
 RA Uytendaele H., Marazzi G., Wu G., Yan Q., Sasseon D., Kitajewski J.;
 RT "Notch4/int-3, a mammary proto-oncogene, is an endothelial
 RT cell-specific mammalian Notch gene."; Development 122:2251-2259(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Rowen L., Mahirae G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
 RA Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
 RT "Sequence of the mouse major histocompatibility locus class III
 RT region."; Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 1436-1600 FROM N.A.
 RX MEDLINE=99252212; PubMed=10233982;
 RA Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
 RT "Intracellular type A particle-mediated activation of the Notch4/int3
 RT gene in a mouse mammary tumor: generation of truncated Notch4/int3
 RT mRNAs by retroviral splicing events."; J. Virol. 73:5166-5171(1999).
 RN [6]
 RP FUNCTION.
 RX MEDLINE=21244657; PubMed=11344305;
 RA Uytendaele H., Ho J., Rossant J., Kitajewski J.;
 RT "Vascular patterning defects associated with expression of activated
 RT Notch4 in embryonic endothelium."; Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648(2001).
 RN [7]
 RP SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
 OF VAL-1463.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeder E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (NL-4) undergo presenilin-dependent
 RT proteolysis."; J. Biol. Chem. 276:40268-40273(2001).
 RN [8]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members."; Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (by similarity). May regulate branching
 CC morphogenesis in the developing vascular system.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytic processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart
 CC kidney, and at lower levels in the ovary and skeletal muscle. A
 CC very low expression is seen in the brain, intestine, liver and
 CC testis.
 CC -1- DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during
 CC embryonic development from 9.0 dpc.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -1- PTM: Phosphorylated.

CC -1- DISEASE: Loss of the extracellular domain causes constitutive
 CC activation of the Notch protein, which leads to hyperproliferation
 CC of glandular epithelial tissues and development of mammary
 CC carcinomas.
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: Contains 29 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M80456; AAB38377.1; -.
 CC EMBL; U43691; AAC52630.1; -.
 CC EMBL; U43691; AAC52631.1; -.
 CC EMBL; AF030001; AAB82004.1; -.
 CC EMBL; AB016771; BAA32281.1; ALT SEQ.
 CC EMBL; AB016772; BAA32283.1; ALT_INIT.
 CC EMBL; AB016773; BAA32284.1; ALT_INIT.
 CC EMBL; AB016774; BAA32285.1; -.
 CC PIR; T09059; T09059.
 CC HSSP; P08709; 1BR9.
 CC MGD; MG1:107471; Notch4.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR001552; Asx_hydroxyl.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR001438; EGF_IT.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR02049; Laminin_EGF.
 CC InterPro; IPR008000; Notch.
 CC Pfam; PF00023; ank; 6.
 CC Pfam; PF00008; EGF; 27.
 CC Pfam; PF00066; notch; 2.
 CC PRINTS; PR00010; EGFBLD.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC PRINTS; PR01452; NOTCH.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00179; EGF_CA; 11.
 CC SMART; SM00004; NL_2.
 CC PROSITE; PS50088; ANK_REPEAT; 5.
 CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 CC PROSITE; PS00010; ASX_HYDROXYL; 11.
 CC PROSITE; PS00022; EGF_1; 28.
 CC PROSITE; PS01186; EGF_2; 21.
 CC PROSITE; PS01187; EGF_CA; 9.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.
 CC SIGNAL 1 20
 CC CHAIN 21 1964 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
 CC CHAIN 1411 1964 TRANSFORMING PROTEIN INT-3.
 CC CHAIN 1428 1964 NOTCH INTRACELLULAR TRUNCATION.
 CC CHAIN 1463 1964 NOTCH INTRACELLULAR DOMAIN.
 CC TRANSNEM 21 1443 EXTRACTED FROM (POTENTIAL).
 CC TRANSNEM 1444 1464 POTENTIAL.
 CC TRANSNEM 1465 1964 CYTOPLASMIC (POTENTIAL).
 CC DOMAIN 21 60 EGF-LIKE 1.
 CC DOMAIN 61 112 EGF-LIKE 2.
 CC DOMAIN 115 152 EGF-LIKE 3.
 CC DOMAIN 153 189 EGF-LIKE 4.
 CC DOMAIN 191 229 EGF-LIKE 5.
 CC DOMAIN 231 271 EGF-LIKE 6.
 CC DOMAIN 273 309 EGF-LIKE 7.
 CC DOMAIN 311 350 EGF-LIKE 8.
 CC DOMAIN 352 388 EGF-LIKE 9.
 CC DOMAIN 389 427 EGF-LIKE 10.
 CC DOMAIN 429 470 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 472 508 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 510 546 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 548 584 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 586 622 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 623 656 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 658 686 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 688 724 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 726 762 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 764 800 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 803 839 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 841 877 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 878 924 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 926 962 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 964 1000 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1002 1040 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1042 1081 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1083 1122 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1126 1167 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
 FT REPEAT 1168 1208 LIN/NOTCH 1.
 FT REPEAT 1209 1242 LIN/NOTCH 2.
 FT REPEAT 1243 1282 LIN/NOTCH 3.
 FT REPEAT 1628 1657 ANK 1.
 FT REPEAT 1661 1691 ANK 2.
 FT REPEAT 1695 1724 ANK 3.
 FT REPEAT 1728 1757 ANK 4.
 Query Match 1.6%; Score 8; DB 1; Length 1964;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 390 GTTPMLA 397
 Db 1696 GTTPMLA 1703
 RESULT 20
 NTCA HUMAN STANDARD; PRT; 2003 AA.
 AC 099456; 000306; Q99458; Q99459; Q99460; Q99461; Q99462; Q99463; Q99464; Q99465; Q99466; Q99467; Q99468; Q99469; Q99470; Q99471; Q99472; Q99473; Q99474; Q99475; Q99476; Q99477; Q99478; Q99479; Q99480; Q99481; Q99482; Q99483; Q99484; Q99485; Q99486; Q99487; Q99488; Q99489; Q99490; Q99491; Q99492; Q99493; Q99494; Q99495; Q99496; Q99497; Q99498; Q99499; Q99500; Q99501; Q99502; Q99503; Q99504; Q99505; Q99506; Q99507; Q99508; Q99509; Q99510; Q99511; Q99512; Q99513; Q99514; Q99515; Q99516; Q99517; Q99518; Q99519; Q99520; Q99521; Q99522; Q99523; Q99524; Q99525; Q99526; Q99527; Q99528; Q99529; Q99530; Q99531; Q99532; Q99533; Q99534; Q99535; Q99536; Q99537; Q99538; Q99539; Q99540; Q99541; Q99542; Q99543; Q99544; Q99545; Q99546; Q99547; Q99548; Q99549; Q99550; Q99551; Q99552; Q99553; Q99554; Q99555; Q99556; Q99557; Q99558; Q99559; Q99560; Q99561; Q99562; Q99563; Q99564; Q99565; Q99566; Q99567; Q99568; Q99569; Q99570; Q99571; Q99572; Q99573; Q99574; Q99575; Q99576; Q99577; Q99578; Q99579; Q99580; Q99581; Q99582; Q99583; Q99584; Q99585; Q99586; Q99587; Q99588; Q99589; Q99590; Q99591; Q99592; Q99593; Q99594; Q99595; Q99596; Q99597; Q99598; Q99599; Q99600; Q99601; Q99602; Q99603; Q99604; Q99605; Q99606; Q99607; Q99608; Q99609; Q99610; Q99611; Q99612; Q99613; Q99614; Q99615; Q99616; Q99617; Q99618; Q99619; Q99620; Q99621; Q99622; Q99623; Q99624; Q99625; Q99626; Q99627; Q99628; Q99629; Q99630; Q99631; Q99632; Q99633; Q99634; Q99635; Q99636; Q99637; Q99638; Q99639; Q99640; Q99641; Q99642; Q99643; Q99644; Q99645; Q99646; Q99647; Q99648; Q99649; Q99650; Q99651; Q99652; Q99653; Q99654; Q99655; Q99656; Q99657; Q99658; Q99659; Q99660; Q99661; Q99662; Q99663; Q99664; Q99665; Q99666; Q99667; Q99668; Q99669; Q99670; Q99671; Q99672; Q99673; Q99674; Q99675; Q99676; Q99677; Q99678; Q99679; Q99680; Q99681; Q99682; Q99683; Q99684; Q99685; Q99686; Q99687; Q99688; Q99689; Q99690; Q99691; Q99692; Q99693; Q99694; Q99695; Q99696; Q99697; Q99698; Q99699; Q99700; Q99701; Q99702; Q99703; Q99704; Q99705; Q99706; Q99707; Q99708; Q99709; Q99710; Q99711; Q99712; Q99713; Q99714; Q99715; Q99716; Q99717; Q99718; Q99719; Q99720; Q99721; Q99722; Q99723; Q99724; Q99725; Q99726; Q99727; Q99728; Q99729; Q99730; Q99731; Q99732; Q99733; Q99734; Q99735; Q99736; Q99737; Q99738; Q99739; Q99740; Q99741; Q99742; Q99743; Q99744; Q99745; Q99746; Q99747; Q99748; Q99749; Q99750; Q99751; Q99752; Q99753; Q99754; Q99755; Q99756; Q99757; Q99758; Q99759; Q99760; Q99761; Q99762; Q99763; Q99764; Q99765; Q99766; Q99767; Q99768; Q99769; Q99770; Q99771; Q99772; Q99773; Q99774; Q99775; Q99776; Q99777; Q99778; Q99779; Q99780; Q99781; Q99782; Q99783; Q99784; Q99785; Q99786; Q99787; Q99788; Q99789; Q99790; Q99791; Q99792; Q99793; Q99794; Q99795; Q99796; Q99797; Q99798; Q99799; Q99800; Q99801; Q99802; Q99803; Q99804; Q99805; Q99806; Q99807; Q99808; Q99809; Q99810; Q99811; Q99812; Q99813; Q99814; Q99815; Q99816; Q99817; Q99818; Q99819; Q99820; Q99821; Q99822; Q99823; Q99824; Q99825; Q99826; Q99827; Q99828; Q99829; Q99830; Q99831; Q99832; Q99833; Q99834; Q99835; Q99836; Q99837; Q99838; Q99839; Q99840; Q99841; Q99842; Q99843; Q99844; Q99845; Q99846; Q99847; Q99848; Q99849; Q99850; Q99851; Q99852; Q99853; Q99854; Q99855; Q99856; Q99857; Q99858; Q99859; Q99860; Q99861; Q99862; Q99863; Q99864; Q99865; Q99866; Q99867; Q99868; Q99869; Q99870; Q99871; Q99872; Q99873; Q99874; Q99875; Q99876; Q99877; Q99878; Q99879; Q99880; Q99881; Q99882; Q99883; Q99884; Q99885; Q99886; Q99887; Q99888; Q99889; Q99890; Q99891; Q99892; Q99893; Q99894; Q99895; Q99896; Q99897; Q99898; Q99899; Q99900; Q99901; Q99902; Q99903; Q99904; Q99905; Q99906; Q99907; Q99908; Q99909; Q99910; Q99911; Q99912; Q99913; Q99914; Q99915; Q99916; Q99917; Q99918; Q99919; Q99920; Q99921; Q99922; Q99923; Q99924; Q99925; Q99926; Q99927; Q99928; Q99929; Q99930; Q99931; Q99932; Q99933; Q99934; Q99935; Q99936; Q99937; Q99938; Q99939; Q99940; Q99941; Q99942; Q99943; Q99944; Q99945; Q99946; Q99947; Q99948; Q99949; Q99950; Q99951; Q99952; Q99953; Q99954; Q99955; Q99956; Q99957; Q99958; Q99959; Q99960; Q99961; Q99962; Q99963; Q99964; Q99965; Q99966; Q99967; Q99968; Q99969; Q99970; Q99971; Q99972; Q99973; Q99974; Q99975; Q99976; Q99977; Q99978; Q99979; Q99980; Q99981; Q99982; Q99983; Q99984; Q99985; Q99986; Q99987; Q99988; Q99989; Q99990; Q99991; Q99992; Q99993; Q99994; Q99995; Q99996; Q99997; Q99998; Q99999; Q100000.
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 4 precursor (Notch 4)
 DE (Notch4).
 GN NOTCH4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND POLYMORPHISM OF POLY-LEU.
 RC TISSUE=Placenta;
 RX MEDLINE=97311416; PubMed=9168133;
 RA Sugaya K., Sasamura S.-I., Nohata J., Kimura T., Fukagawa T.,
 RA Nakamura Y., Ando A., Inoko H., Ikemura T., Mita K.;
 RT "Gene organization of human NOTCH4 and (C)Gn polymorphism in this
 RT human counterpart gene of mouse proto-oncogene Int3.";
 RL Gene 189:235-244(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RC TISSUE=Bone marrow, and Heart;
 RX MEDLINE=98360091; PubMed=9693032;
 RA Li L., Huang G.M., Banta A.B., Deng Y., Smith T., Dong P.,
 RA Friedman C., Chen L., Task B.J., Spies T., Rowen L., Hood L.;
 RT "Cloning, characterization, and the complete 56.8-kilobase DNA
 RT sequence of the human NOTCH4 gene.";
 RL Genomics 51:45-58(1998).
 RN [3]
 RP SEQUENCE OF 1-503 FROM N.A., AND VARIANTS GLN-117 AND GLN-317.
 RA Miyagawa T., Tokunaga K., Hojho H.;
 RT "Human notch4 gene variant.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]


```

FT REPEAT 1733 1765 ANK 4.
FT REPEAT 1766 1798 ANK 5.
FT DISULFID 28 41 BY SIMILARITY.
FT DISULFID 35 51 BY SIMILARITY.
FT DISULFID 53 62 BY SIMILARITY.
FT DISULFID 68 80 BY SIMILARITY.
FT DISULFID 74 103 BY SIMILARITY.
FT DISULFID 105 114 BY SIMILARITY.
FT DISULFID 122 133 BY SIMILARITY.
FT DISULFID 127 143 BY SIMILARITY.
FT DISULFID 145 154 BY SIMILARITY.
FT DISULFID 160 171 BY SIMILARITY.
FT DISULFID 165 180 BY SIMILARITY.
FT DISULFID 182 191 BY SIMILARITY.
FT DISULFID 198 211 BY SIMILARITY.
FT DISULFID 205 220 BY SIMILARITY.
FT DISULFID 222 231 BY SIMILARITY.
FT DISULFID 238 249 BY SIMILARITY.

Query Match
Best Local Similarity 1.6%; Score 8; DB 1; Length 2003;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 390 GTTPLMLA 397
Db 1701 GTTPLMLA 1708

RESULT 21
PBTG_CVACA STANDARD; PRT: 35 AA.
ID AC Opt03;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome b6-f complex subunit V (Cytochrome b6f complex subunit
DE pbtg).
GN PBTG.
OS Cyanidium caldarium.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC Cyanidium.
OX NCBI_TaxID=2771;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RK-1;
RX MEDLINE=20496959; PubMed=11040290;
RA Gloeocner G., Rosenthal A., Valentin K.-U.;
RT "The structure and gene repertoire of an ancient red algal plastid
RT genome."
RL J. Mol. Evol. 51:382-390(2000).
CC -1- FUNCTION: The cytochrome b6-f complex functions in the linear
CC cross-membrane transport of electrons between photosystem II and
CC I, as well as in cyclic electron flow around photosystem I. Pbtg
CC is required for either the stability or assembly of the cytochrome
CC b6-f complex (By similarity).
CC -1- SUBCELLULAR LOCATION: Thylakoid membrane-associated.
CC -1- SIMILARITY: Belongs to the pbtg family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF022186; AAF12884.1; -
CC HAMAP: MF_00432; -; 1.
CC InterPro: IPR003683; Cytochrome_b6_f_5.
CC Pfam: PF02529; Pbtg, 1.
CC Electon transport; Chloroplast; Respiratory chain; Thylakoid;
CC Transmembrane.

```

```

FT DOMAIN 1 4 LUMENAL (POTENTIAL).
FT TRANSMEM 5 25 POTENTIAL.
FT DOMAIN 26 35 STROMAL (POTENTIAL).
SQ SEQUENCE 35 AA; 3803 MW; B03C27094A1B74F3 CRC64;

Query Match
Best Local Similarity 1.4%; Score 7; DB 1; Length 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 365 VLTGIV 371
Db 4 VLTGIV 10

RESULT 22
TX31_DENAN STANDARD; PRT: 80 AA.
ID AC P18329;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Toxin C13S1C1 precursor.
OS Dendroaspis angusticeps (Eastern green mamba).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Elapidae; Elapinae; Dendroaspis.
OX NCBI_TaxID=8618;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA Duncanel F.;
RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 22-80.
RC TISSUE=Venom;
RX MEDLINE=80242820; PubMed=7397227;
RA Joubert F.J., Taljaard N.;
RT "The complete primary structures of two reduced and
RT S-carboxymethylated angusticeps-type toxins from Dendroaspis
RT angusticeps (green mamba) venom."
RL Blochim. Biophys. Acta 623:449-456(1980).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -1- MISCELLANEOUS: LD(50) IS 40 +- 3 MG/KG BY INTRAVENOUS INJECTION.
CC -1- SIMILARITY: Belongs to the snake toxin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X52260; CAA36503.1; -
CC EMBL: X53410; CAA37486.1; -
CC PIR: S10464; S10464.
CC HSBP; P01416; INTX.
CC InterPro: IPR003571; Snake_toxin.
CC Pfam: PF00087; toxin, 1.
CC ProDom: PD000206; Snake_toxin, 1.
CC PROSITE: PS00272; SNAKE_TOXIN, 1.
KM Toxin; Signal.
FT SIGNAL 1 21
FT CHAIN 22 80 TOXIN C13S1C1.
FT DISULFID 24 43 BY SIMILARITY.
FT DISULFID 38 60 BY SIMILARITY.
FT DISULFID 62 73 BY SIMILARITY.
FT DISULFID 74 79 BY SIMILARITY.
FT CONFLICT 56 56 MISSING (IN REF. 2).
SQ SEQUENCE 80 AA; 8954 MW; 5F4450DF19E721C8 CRC64;

Query Match
1.4%; Score 7; DB 1; Length 80;

```


Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 288 LLLTLTV 294
DB 4 LLLTLTV 10

RESULT 23

TXF8_DENAN STANDARD; PRT; 81 AA.
ID TXF8_DENAN
AC P01404;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Toxin F-VIII precursor (Toxin TA2) (Toxin DBF8).
OS Dendroaspis angusticeps (Eastern green mamba).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Elapidae; Elapinae; Dendroaspis.
OX NCBI_TaxID=8618;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA Ducancel F.;
RL Submitted (JUN-1990) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 22-81.
RC TISSUE=Venom;
RA MEDLINE=74070661; PubMed=4203470;
RA Viljoen C.C., Botes D.P.;
RT "Snake venom toxins. The purification and amino acid sequence of
RT toxin TA2 from Dendroaspis angusticeps venom.";
RL J. Biol. Chem. 249:366-372(1974).
CC -1- FUNCTION: Its pharmacological action is unknown. It probably acts
CC synergistically with other venom components as whole venom is very
CC toxic.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -1- MISCELLANEOUS: LD(50) is >250 mg/kg by subcutaneous injection.
CC -1- SIMILARITY: Belongs to the snake toxin family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X53409; CAA37485.1; -.
DR PIR; S21299; T5EP2A.
DR HSSP; P01416; INTX.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Toxin; Signal.
FT SIGNAL 1 21
FT CHAIN 22 81
FT DISULFID 24 43
FT DISULFID 38 60
FT DISULFID 62 73
FT DISULFID 74 79
FT CONFLICT 54 54
SQ SEQUENCE 81 AA; 8849 MW; 733195E17A656C97 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 81;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 288 LLLTLTV 294
DB 4 LLLTLTV 10

DB 4 LLLTLTV 10

RESULT 24

CBTB_NAJAT STANDARD; PRT; 82 AA.
ID CBTB_NAJAT
AC P80558; Q42285;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cobrotoxin b precursor (CBT b) (Short neurotoxin).
OS Naja atra (Chinese cobra).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Elapidae; Elapinae; Naja.
OX NCBI_TaxID=8656;
RN [1]
RP SEQUENCE FROM N.A., AND REVISION TO 67.
RC TISSUE=Liver, and Venom;
RA MEDLINE=98158338; PubMed=9498573;
RA Chang L.-S., Chou Y.-C., Lin S.-R., Wu B.-N., Lin J., Hong E.,
RA Sun Y.-J., Hsiao C.-D.;
RT "A novel neurotoxin, cobrotoxin b, from Naja atra (Taiwan cobra)
RT venom: purification, characterization, and gene organization.";
RL J. Biochem. 122:1252-1259(1997).
RN [2]
RP SEQUENCE FROM N.A., AND REVISION TO 67.
RC TISSUE=Venom gland;
RA Chu R.C., Yang C.C.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Produces peripheral paralysis by blocking neuromuscular
CC transmission at the postsynaptic site. Binds to the nicotinic
CC acetylcholine receptor (by similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -1- SIMILARITY: Belongs to the snake toxin family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Y13399; CAA73829.2; -.
DR EMBL; AF031472; AAB86636.1; -.
DR PIR; JCS892; JCS892.
DR HSSP; P01427; INOR.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Toxin; Neurotoxin; Postsynaptic neurotoxin; Signal.
FT SIGNAL 1 21
FT CHAIN 22 82
FT DISULFID 24 44
FT DISULFID 38 61
FT DISULFID 63 74
FT DISULFID 75 80
SQ SEQUENCE 82 AA; 9139 MW; 1FPA21189C08BEB8 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 288 LLLTLTV 294
DB 4 LLLTLTV 10

RESULT 25
CBTA_NAJAT

ID CHTA NAJAT STANDARD; PRT; 83 AA.
AC P01430; 013079; Created
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cobrotoxin precursor (CBT) (Short neurotoxin 1).
OS Naja atra (Chinese cobra), and
OS Naja kaouthia (Monocled cobra) (Naja siamensis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodactylia; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Elapinae; Naja.
OX NCBI_TaxID=8656, 8649;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=N.atra; TISSUE=Venom gland;
RA Chu C.C.; Yang C.C.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=N.atra; TISSUE=Liver;
RX MEDLINE=9804246; PubMed=9367842;
RA Chang L.-S., Lin J., Chou Y.-C., Hong E.;
RT "Genomic structures of cardiotoxin 4 and cobrotoxin from Naja naja atra (Taiwan cobra).";
RL Biochem. Biophys. Res. Commun. 239:756-762(1997).
RN [3]
RP SEQUENCE OF 22-83.
RC SPECIES=N.atra; TISSUE=Venom;
RX MEDLINE=69294410; PubMed=5820687;
RA Yang C.C., Yang H.J., Huang J.S.;
RT "The amino acid sequence of cobrotoxin.";
RL Biochim. Biophys. Acta 188:65-77(1969).
RN [4]
RP SEQUENCE OF 22-83.
RC SPECIES=N.atra; TISSUE=Venom;
RX MEDLINE=93192268; PubMed=8448165;
RA Chiu S.H., Raynor R.L., Zheng B., Chambers T.C., Kuo J.F.;
RT "Cobra venom cardiotoxin (cytotoxin) isoforms and neurotoxin: comparative potency of protein kinase C inhibition and cancer cell cytotoxicity and modes of enzyme inhibition.";
RL Biochemistry 32:2062-2067(1993).
RN [5]
RP SEQUENCE OF 22-83.
RC SPECIES=N.kaouthia; TISSUE=Venom;
RX MEDLINE=22035208; PubMed=12039691;
RA Meng Q.-X., Wang W.-Y., Lu Q.-M., Jin Y., Wei J.-F., Zhu S.-W., Xiong Y.-L.;
RT "A novel short neurotoxin, cobrotoxin C, from monocellate cobra (Naja kaouthia) venom: isolation and purification, primary and secondary structure determination, and tertiary structure modeling.";
RL Comp. Biochem. Physiol. 132C:113-121(2002).
RN [6]
RP DISULFIDE BONDS.
RC SPECIES=N.atra;
RX MEDLINE=71125117; PubMed=5533659;
RA Yang C.C., Yang H.J., Chiu R.H.C.;
RT "The position of disulfide bonds in cobrotoxin.";
RL Biochim. Biophys. Acta 214:355-363(1970).
RN [7]
RP STRUCTURE BY NMR.
RC SPECIES=N.atra;
RX MEDLINE=91065386; PubMed=2249693;
RA Yu C., Lee C.-S., Chuang L.-C., Shet Y.-R., Wang C.Y.;
RT "Two-dimensional NMR studies and secondary structure of cobrotoxin in aqueous solution.";
RL Eur. J. Biochem. 193:789-799(1990).
RN [8]
RP STRUCTURE BY NMR.
RC SPECIES=N.atra;
RX MEDLINE=93183875; PubMed=8443154;
RA Yu C., Bhaskaran R., Chuang L.-C., Yang C.C.;
RT "Solution conformation of cobrotoxin: a nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing study.";

RL Biochemistry 32:2131-2136(1993).
CC -1- FUNCTION: Produces peripheral paralysis by blocking neuromuscular transmission at the postsynaptic site. Binds to the nicotinic acetylcholine receptor.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -1- MISCELLANEOUS: LD(50) is 0.09 mg/kg by subcutaneous injection.
CC -1- SIMILARITY: Belongs to the snake toxin family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; U58519; AAB03221.1; -;
DR EMBL; U58520; AAB03222.1; -;
DR EMBL; U58521; AAB03223.1; -;
DR EMBL; U77490; AAB36930.1; -;
DR EMBL; U77491; AAB36931.1; -;
DR EMBL; U77492; AAB36932.1; -;
DR EMBL; U42582; AAB01538.1; -;
DR EMBL; Y12492; CAA73097.2; -;
DR PIR; J05769; NINJIF.
DR PDB; 1COD; 20-JUL-95.
DR PDB; 1COE; 20-JUL-95.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; Toxin, 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KM Toxin; Neurotoxin; Postsynaptic neurotoxin; Multigene family; Signal;
KM 3D-structure.
FT SIGNAL 1 21
FT CHAIN 22 83 COBROTOXIN.
FT DISULFID 24 45
FT DISULFID 38 62
FT DISULFID 64 75
FT DISULFID 76 81
FT STRAND 24 27
FT STRAND 33 36
FT STRAND 45 50
FT TURN 53 54
FT STRAND 57 59
FT STRAND 62 62
FT STRAND 72 76
SQ SEQUENCE 83 AA; 9261 MW; 4DD6077C92717052 CRC64;
Query March 1.4%; Score 7; DB 1; Length 83;
Best local similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 288 LLTTLV 294
DB 4 LLTTLV 10
RESULT 26
HG15 CHICK STANDARD; PRT; 104 AA.
AC P12902;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Nonhistone chromosomal protein HMG-14A.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=88255874; PubMed=3384337;
RA Dodgson J.B., Browne D.L., Black A.J.;
RT "Chicken chromosomal protein HMG-14 and HMG-17 cDNA clones:
RT isolation, characterization and sequence comparison.";
RL Gene 63:287-295(1988).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=White leghorn;
RX MEDLINE=93185924; PubMed=8444343;
RA Browne D.L., Dodgson J.B.;
RT "The gene encoding chicken chromosomal protein HMG-14a is transcribed
RT into multiple mRNAs.";
RL Gene 124:199-206(1993).
CC
CC -1- FUNCTION: BIND TO THE INNER SIDE OF THE NUCLEOSOMAL DNA THUS
CC ALTERING THE INTERACTION BETWEEN THE DNA AND THE HISTONE OCTAMER.
CC MAY BE INVOLVED IN THE PROCESS WHICH MAINTAINS TRANSCRIBABLE GENES
CC IN A UNIQUE CHROMATIN CONFORMATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- MISCELLANEOUS: THERE ARE TWO HMG-14 PROTEINS IN CHICKEN: HMG-14A
CC -1- THE MAJOR COMPONENT, AND HMG-14B THE MINOR COMPONENT.
CC -1- SIMILARITY: BELONGS TO THE HMG-14/HMG-17 PROTEIN FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M26675; AAA48815.1; -
DR EMBL; X63083; CAA44804.1; -
DR EMBL; X63084; CAA44804.1; JOINED.
DR EMBL; X63085; CAA44804.1; JOINED.
DR EMBL; X63086; CAA44804.1; JOINED.
DR EMBL; X63087; CAA44804.1; JOINED.
DR PIR; S22122; NSCHM4.
DR InterPro: IPR000079; HMG14_17.
DR Pfam; PF01101; HMG14_17.1.
DR PRINTS; PRO0925; NONHISHMG17.
DR SMART; SM00527; HMG17; 1.
DR PROSITE; PS00355; HMG14_17.1.
KW Nuclear protein; DNA-binding; Multigene family.
FT INIT MET 0
SQ SEQUENCE 104 AA; 1125 MW; SEC6F5F165611316 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 104;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 KKKKGAK 269
DB 58 KKKKGAK 64

RESULT 27
RL3E_SULSO STANDARD; PRT; 108 AA.
AC Q980R3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L30e.
GN RPL30E OR SSO0221.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=1142726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,

```

```

RA Aways M.J., Chan-Wehner C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Tni-Ygoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC
CC -1- SIMILARITY: BELONGS TO THE L30E FAMILY OF RIBOSOMAL PROTEINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE006659; AAK40564.1; -
DR PIR; P90163; E90163.
DR HAMAP; MF_00481; 1.
DR InterPro: IPR000231; Ribosomal_L30e.
DR InterPro: IPR004038; Ribosomal_L7A.
DR Pfam; PF01248; Ribosomal_L7ae; 1.
DR ProDom; PD004495; Ribosomal_L30e; 1.
DR PROSITE; PS00709; RIBOSOMAL_L30E_1.
DR PROSITE; PS00993; RIBOSOMAL_L30E_2; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 108 AA; 12095 MW; 4B24861EC334AA2C CRC64;

Query Match 1.4%; Score 7; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 YAKLSNI 88
DB 60 YAKLSNI 66

RESULT 28
RL3E_METUA STANDARD; PRT; 110 AA.
AC P54051;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L30e.
GN RPL30E OR MJ1044.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake O., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utermack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurt M.A., Kaine B.F., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC
CC -1- SIMILARITY: BELONGS TO THE L30E FAMILY OF RIBOSOMAL PROTEINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U67547; AAB99048.1; -
 DR PIR; C64430; C64430.
 DR HSSP; P14120; 1CN9.
 DR TIGR; M1044; -
 DR HAMAP; MF_00481; -; 1.
 DR InterPro; IPR000231; Ribosomal_L30e.
 DR InterPro; IPR004038; Ribosomal_L7A.
 DR Pfam; PF01248; Ribosomal_L7Ae; 1.
 DR ProDom; PD004495; Ribosomal_L30e; 1.
 DR PROSITE; PS00709; RIBOSOMAL_L30E_1; 1.
 DR PROSITE; PS00993; RIBOSOMAL_L30E_2; 1.
 KM Ribosomal protein; Complete proteome.
 SQ SEQUENCE 110 AA; 12134 MW; EBBCA4E135E098B6 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 110;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 YAKLSNI 88
 |||||
 Db 58 YAKLSNI 64

RESULT 29
 HBBO_PAGBO STANDARD; PRT; 146 AA.
 ID_HBBO_PAGBO
 AC P82345;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hemoglobin beta-0 chain (HB 0).
 GN HBBO.
 OS Pagothenia borchgrevinki (Bald rockcod) (Trematomus borchgrevinki).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostei;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciforma; Perciformes;
 OC Nototheniidae; Nototheniidae; Pagothenia.
 OC NCBI_TaxID=8213;
 RN [1]
 RP SEQUENCE.
 RC Tissue-Blood;
 RA Riccio A., Tamburrini M., Carratore V., di Prisco G.;
 RT "Functionally distinct haemoglobins of the cryopelagic antarctic teleost Pagothenia borchgrevinki.";
 RL J. Fish Biol. 57:20-32(2000).
 CC -1- FUNCTION: INVOLVED IN OXYGEN TRANSPORT FROM THE LUNG TO THE VARIOUS PERIPHERAL TISSUES.
 CC -1- SUBUNIT: Heterotetramer of two alpha chains and two beta chains.
 CC -1- TISSUE SPECIFICITY: Red blood cells.
 CC -1- MISCELLANEOUS: THIS FISH HAS FIVE HEMOGLOBINS: HB C, HB O, HB 1, HB 2 AND HB 3. HB 0 PRESENTS THE STRONGEST BOHR EFFECT WHILE HB 1 PRESENTS THE WEAKEST BOHR EFFECT.
 CC -1- SIMILARITY: Belongs to the globin family.
 DR HSSP; P80044; 1PBX.
 DR InterPro; IPR000971; Globin.
 DR Pfam; PF00042; globin; 1.
 DR PROSITE; PS01033; GLOBIN; 1.
 KM Heme; Oxygen transport; Transport; Erythrocyte.
 FT METAL 63 63 IRON (HEME DISTAL LIGAND).
 FT METAL 92 92 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 146 AA; 16364 MW; 1DC991617574CF73 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 146;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 441 VVAARFG 447
 |||||
 Db 113 VVAARFG 119

RESULT 30
 HBBC_PAGBE STANDARD; PRT; 146 AA.
 ID_HBBC_PAGBE
 AC P45722;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemoglobin beta-C chain.
 GN HBBC.
 OS Pagothenia bernacchi (Emerald rockcod) (Trematomus bernacchi).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostei;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciforma; Perciformes;
 OC Nototheniidae; Nototheniidae; Trematomus.
 OC NCBI_TaxID=40690;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=94193650; PubMed=8144556;
 RA D'Avino R., Caruso C., Tamburrini M., Romano M., Rutigliano B., Polverino de Laureto P., Camardella L., Carratore V., di Prisco G.;
 RT "Molecular characterization of the functionally distinct hemoglobins of the Antarctic fish Trematomus newnesi.";
 RL J. Biol. Chem. 269:9675-9681(1994).
 CC -1- FUNCTION: COMPONENT OF HEMOGLOBIN C, A MINOR GLOBIN OF THIS FISH.
 CC -1- SUBUNIT: HEMOGLOBIN C (HB C) IS A TETRAMER OF TWO ALPHA CHAINS AND TWO BETA-C CHAINS.
 CC -1- SIMILARITY: Belongs to the globin family.
 DR HSSP; P80044; 1PBX.
 DR InterPro; IPR000971; Globin.
 DR Pfam; PF00042; globin; 1.
 DR PROSITE; PS01033; GLOBIN; 1.
 KM Heme; Oxygen transport; Transport; Erythrocyte.
 FT METAL 63 63 IRON (HEME DISTAL LIGAND).
 FT METAL 92 92 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 146 AA; 16351 MW; 990E4DA62AF046 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 146;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 441 VVAARFG 447
 |||||
 Db 113 VVAARFG 119

RESULT 31
 HBBC_TRENE STANDARD; PRT; 146 AA.
 ID_HBBC_TRENE
 AC P45721;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hemoglobin beta-C chain.
 GN Trematomus newnesi (Dusky notothen).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciforma; Perciformes;
 OC Nototheniidae; Nototheniidae; Trematomus.
 OC NCBI_TaxID=35730;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=94193650; PubMed=8144556;
 RA D'Avino R., Caruso C., Tamburrini M., Romano M., Rutigliano B., Polverino de Laureto P., Camardella L., Carratore V., di Prisco G.;
 RT "Molecular characterization of the functionally distinct hemoglobins of the Antarctic fish Trematomus newnesi.";
 RL J. Biol. Chem. 269:9675-9681(1994).
 CC -1- FUNCTION: COMPONENT OF HEMOGLOBIN C, A MINOR (20% TO 25%) GLOBIN OF THIS FISH.
 CC -1- SUBUNIT: HEMOGLOBIN C (HB C) IS A TETRAMER OF TWO ALPHA-1

CC CHAINS AND TWO BETA-C CHAINS.
 CC -1- SIMILARITY: Belongs to the globin family.
 DR PIR, D54403, D54403.
 DR HSSB, P80044, 1PBX.
 DR InterPro, IPR002337, Beta_haem.
 DR InterPro, IPR000971, Globin.
 DR Pfam, PF00042, Globin, 1.
 DR PRINTS, PR00814, BETAHAEM.
 DR PROSITE, PS01033, GLOBIN, 1.
 KM Heme; Oxygen transport; Transport; Erythrocyte.
 FT METAL 63 63 IRON (HEME DISTAL LIGAND).
 FT METAL 92 92 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 146 AA; 16260 MW; 9C05E8B6E092979 CRC64;

Query March 1.4%; Score 7; DB 1; Length 146;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 441 VVAARFG 447
 |||||
 Db 113 VVAARFG 119

RESULT 32
 NU6M LOBL STANDARD; PRT; 168 AA.
 AC 047478;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3).
 GN Nde.
 OS Loligo bleekeri (Bleeker's squid).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
 OC Decapodiformes; Loliginidae; Loligo.
 OX NCBI_TaxID=6617;
 RN [1]
 RA SEQUENCE FROM N.A.
 RA Tomita K., Ueda T., Watanabe K.;
 RT "Completing of squid (Loligo bleekeri) mitochondrial genome
 sequencing.";
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB009838; BAA24061.1; -
 DR EMBL; AB029616; BAB03648.1; -
 DR InterPro, IPR001457; Oxidored_q3.
 DR Pfam, PF00499; oxidored_q3; 1.
 KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
 SQ SEQUENCE 168 AA; 19301 MW; DBA1A1726BF5CBA8 CRC64;

Query March 1.4%; Score 7; DB 1; Length 168;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 472 LAVILLF 478
 |||||
 Db 140 LAVILLF 146

RESULT 33
 NU6M CYPCA STANDARD; PRT; 172 AA.
 AC P24982;

DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3).
 GN MTD6 OR ND6.
 OS Cyprinus carpio (Common carp).
 OC Cyprinus carpio (Common carp).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Cyprinus.
 OX NCBI_TaxID=7962;
 RN [1]
 RA SEQUENCE FROM N.A.
 RP MEDLINE=94223691; PubMed=8169959;
 RX Chang Y.S., Huang P.L., Lo T.B.;
 RT "The complete nucleotide sequence and gene organization of carp
 (Cyprinus carpio) mitochondrial genome.";
 RL J. Mol. Evol. 38:138-155(1994).
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X61010; CAA43338.1; -
 DR EMBL; X61010; CAA43338.1; -
 DR PIR, S36007; S36007.
 DR InterPro, IPR001457; Oxidored_q3.
 DR Pfam, PF00499; oxidored_q3; 1.
 KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
 SQ SEQUENCE 172 AA; 17971 MW; B3EBADA8B8E588 CRC64;

Query March 1.4%; Score 7; DB 1; Length 172;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 288 LLTLTLV 294
 |||||
 Db 150 LLTLTLV 156

RESULT 34
 NU6M CARAU STANDARD; PRT; 173 AA.
 AC 078689;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3).
 GN MTD6 OR ND6.
 OS Carassius auratus (Goldfish).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Carassius.
 OX NCBI_TaxID=7957;
 RN [1]
 RA SEQUENCE FROM N.A.
 RP STRAIN=A23 / Langsdorff; TISSUE=Oocyte;
 RC Murakami M., Yamashita Y., Fujitani H.;
 RT "The complete sequence of mitochondrial genome from a gynogenetic
 triploid 'ginsuna' (Carassius auratus langsdorff).";
 RL Zool. Sci. 15:335-337(1998).
 RN [2]
 RA SEQUENCE FROM N.A.
 RP STRAIN=Cv1er1;
 RC Murakami M.;
 RT "Carassius auratus cv1er1 mitochondrial DNA, complete sequence.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB06953; BAA31249.1; -
CC DR EMBL; AB045144; BAB40359.1; -
CC DR InterPro; IPR001457; Oxidored_q3.
CC DR Pfam; PF00499; oxidored_q3; 1.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
CC SEQUENCE 173 AA; 18095 MW; 3FAE4B5AB50C30C CRC64;
QY Query Match 1.4%; Score 7; DB 1; Length 173;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 288 LLLTLV 294
Db 151 LLLTLV 157
-----
RESULT 35
NM6M_GADMO STANDARD; PRT; 173 AA.
ID NM6M_GADMO
AC P55783;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3).
GN MTND6 OR ND6.
OS Gadus morhua (Atlantic cod).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthopterygii; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
OC NCBI_TaxID=8049;
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Norwegian coastal 1;
RX MEDLINE=96414925; PubMed=8817926;
RA Johansen S., Bakke I.;
RT "The complete mitochondrial DNA sequence of Atlantic cod (Gadus
RT morhua); relevance to taxonomic studies among codfishes.";
RL Mol. Mar. Biol. Biotechnol. 5:203-214(1996).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X99772; CA68116.1; -
CC DR EMBL; X99772; CA68116.1; -
CC DR InterPro; IPR001457; Oxidored_q3.
CC DR Pfam; PF00499; oxidored_q3; 1.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
CC SEQUENCE 173 AA; 18090 MW; 7EE5754621B4C7FCB CRC64;
QY Query Match 1.4%; Score 7; DB 1; Length 173;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 288 LLLTLV 294
Db 151 LLLTLV 157

```

```

RESULT 36
RRF_WIGBR STANDARD; PRT; 185 AA.
ID RRF_WIGBR
AC Q8D2G5;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ribosome recycling factor (Ribosome releasing factor) (RRF).
GN RRF OR WIGBR3890.
OS Wigglesworthia glossinidia brevipalpis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Wigglesworthia.
OX NCBI_TaxID=356970;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22297718; PubMed=12219091;
RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
RA Akao S.;
RT "Genome sequence of the endocellular obligate symbiont of tsetse
RT flies, Wigglesworthia glossinidia."
RL Nat. Genet. 32:402-407(2002).
CC -1- FUNCTION: Responsible for the release of ribosomes from messenger
CC RNA at the termination of protein biosynthesis. May increase the
CC efficiency of translation by recycling ribosomes from one round of
CC translation to another (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE RRF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB063522; BAC24535.1; -
CC DR HAMAP; MF_00040; -; 1
CC DR InterPro; IPR002661; RRF.
CC DR Pfam; PF01765; RRF; 1.
CC DR PRODOM; PD004103; RRF; 1.
CC DR TIGRFAMs; TIGR00496; frr; 1.
CC Protein biosynthesis; Complete proteome.
CC SEQUENCE 185 AA; 21650 MW; C71569E7155DECD CRC64;
QY Query Match 1.4%; Score 7; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 KLSNLS 90
Db 53 KLSNLS 59
-----
RESULT 37
RK24_SPTOL STANDARD; PRT; 191 AA.
ID RK24_SPTOL
AC P27683;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE 50S ribosomal protein L24, chloroplast precursor (Cl24).
GN RPL24.
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=91365238; PubMed=1889743;
RA Carol P., Li Y.F., Maché R.;

```

RT "Conservation and evolution of the nucleus-encoded and chloroplast-specific ribosomal proteins in pea and spinach."

RT Gene 103:139-145(1991).

CC -1- SUBCELLULAR LOCATION: Chloroplast.

CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M58522; AAA34042.1; -.

DR PIR: JH0585; JH0585.

DR InterPro: IPR005824; KOW.

DR InterPro: IPR006646; KOW_sub.

DR InterPro: IPR003256; Ribosomal_L24.

DR InterPro: IPR005825; Ribosomal_L24_26.

DR Pfam: PF00467; KOW; 1.

DR ProDom: PD001677; Ribosomal_L24; 1.

DR SMART: SM00739; KOW; 1.

DR TIGRFAMs: TIGR01079; rplX_bact; 1.

DR PROSITE: PS01108; RIBOSOMAL_L24; 1.

KM Ribosomal protein; Chloroplast; Transit peptide.

FT TRANSIT 1 45 CHLOROPLAST.

FT CHAIN 46 191 50S RIBOSOMAL PROTEIN L24.

FT SEQUENCE 191 AA; 21439 MW; 1BE01D8355824E7D CRC64;

QY 500 LKEQEVA 506

Db 138 LKEQEVA 144

Query March 1.4%; Score 7; DB 1; Length 191;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 38

INF_MEIGA STANDARD; PRT; 192 AA.

ID INF_MEIGA

AC P51527;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Interferon precursor.

OS Melaleuca leucadendron (Common turkay).

OS Melaleuca leucadendron; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Meleagrididae; Melagris.

OC NCBI_TaxID=9103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Nicholas;

RX MEDLINE=96079079; PubMed=7494342;

RA Surenh M., Karaca K., Foster D., Sharma J.M.;

RT "Molecular and functional characterization of turkey interferon."

RL J. Virol. 69:8159-8163(1995).

CC -1- SIMILARITY: BELONGS TO THE INTERFERON ALPHA, BETA AND DELTA FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U28140; AAB40029.1; -.

DR InterPro: IPR000471; Interferon_abd.

DR Pfam: PF00143; Interferon; 1.

DR ProDom: PD000550; Interferon_abd; 1.

DR SMART: SM00076; IFabd; 1.

DR PROSITE: PS00252; INTERFERON A B D; 1.

KM Cytochrome; Antiviral; Glycoprotein; Signal.

FT SIGNAL 1 30 POTENTIAL.

FT CHAIN 31 192 INTERFERON.

FT DISULFID 31 128 BY SIMILARITY.

FT DISULFID 67 167 BY SIMILARITY.

FT DISULFID 60 154 POTENTIAL.

FT CARBOHYD 70 70 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 77 77 N-LINKED (GLCNAC...) (POTENTIAL).

SEQ SEQUENCE 192 AA; 21884 MW; 9ED81EF2DC4FF2C3 CRC64;

QY 287 ILTLTL 293

Db 15 ILTLTL 21

Query March 1.4%; Score 7; DB 1; Length 192;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 39

INF_CHICK STANDARD; PRT; 193 AA.

ID INF_CHICK

AC P42165;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Interferon type A1/A2 precursor.

GN IFNA1 AND IFNA2.

OS Gallus gallus (Chicken).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

OC NCBI_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=9435855; PubMed=7521382;

RA Sekellick M.J., Ferrandino A.F., Hopkins D.A., Marcus P.I.;

RT "Chicken interferon gene: cloning, expression, and analysis."

RL J. Interferon Res. 14:71-79(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=White Leghorn; Tissue=Spleen;

RX MEDLINE=96205950; PubMed=8631799;

RA Sick C., Schultz U., Staehel P.;

RT "A family of genes coding for two serologically distinct chicken interferons."

RL J. Biol. Chem. 271:7635-7639(1996).

CC -1- FUNCTION: HAS ANTIVIRAL ACTIVITIES.

CC -1- SUBCELLULAR LOCATION: Secreted (Probable).

CC -1- DEVELOPMENTAL STAGE: APPEARS FIRST AT 3 HOURS POST-INFECTION, INCREASES TO GIVE THE STRONGEST SIGNAL AT ABOUT 9 HOURS AND GRADUALLY WAXES TO ALMOST NOTHING AT 24 HOURS.

CC -1- SIMILARITY: BELONGS TO THE INTERFERON ALPHA, BETA AND DELTA FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U07868; AAB50213.1; -.

DR EMBL: X92477; CAA63215.1; -.

DR EMBL: X92476; CAA63214.1; -.

DR PIR: I50693; I50693.

DR InterPro: IPR000471; Interferon_abd.

DR Pfam: PF00143; Interferon; 1.

DR PRINTS: PR00266; INTERFERONAB.

DR ProDom: PD000550; Interferon_abd; 1.

```

DR SMART; SMO0076; IFabd; 1.
DR PROSITE; PS00252; INTERFERON A B D; 1.
KM Cycokine; Antiviral; Glycoprotein; signal; Multigene family.
FT SIGNAL 1 31
FT CHAIN 32 193
FT DISULFID 32 129
FT DISULFID 61 155
FT DISULFID 68 168
FT CARBOHYD 65 65
FT CARBOHYD 71 71
FT CARBOHYD 108 108
FT CARBOHYD 186 186
FT VARIANT 65 65
SQ SEQUENCE 193 AA; 22116 MW; 35AF74479C450A15 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILTLTLL 293
DB 15 ILTLTLL 21

RESULT 40
INF3 CHICK STANDARD; PRT; 193 AA.
AC Q90872;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Interferon type A3 precursor.
GN IFNA3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauromia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White Leghorn; TISSUE=Spleen;
RX MEDLINE=96205950; PubMed=8631799;
RA Sick C.; Schultz U.; Staeheli P.;
RT "A family of genes coding for two serologically distinct chicken
interferons." 271:7635-7639(1996).
RL J. Biol. Chem. 271:7635-7639(1996).
CC -1- FUNCTION: HAS ANTIVIRAL ACTIVITIES.
CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC -1- SIMILARITY: BELONGS TO THE INTERFERON ALPHA, BETA AND DELTA
FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X92478; CAA63216.1; -
DR InterPro; IPR000471; Interferon_abd.
DR Pfam; PF00143; Interferon; 1.
DR PRINTS; PR00266; INTERFERONAB.
DR PRODOM; PD000550; Interferon_abd; 1.
DR SMART; SMO0076; IFabd; 1.
DR PROSITE; PS00252; INTERFERON A B D; 1.
KM Cycokine; Antiviral; Glycoprotein; signal; Multigene family.
FT SIGNAL 1 31
FT CHAIN 32 193
FT DISULFID 32 129
FT DISULFID 61 155
FT DISULFID 68 168
FT CARBOHYD 65 65
SQ SEQUENCE 193 AA; 22116 MW; 35AF74479C450A15 CRC64;

```

```

FT CARBOHYD 71 71
FT CARBOHYD 108 108
FT CARBOHYD 186 186
SQ SEQUENCE 193 AA; 22116 MW; B53375279C530B7B CRC64;

Query Match 1.4%; Score 7; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILTLTLL 293
DB 15 ILTLTLL 21

Search completed: November 25, 2003, 10:05:47
Job time : 23 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 09:32:01 ; Search time 64 Seconds
(without alignments)
2076.517 Million cell updates/sec

Title: US-09-869-433-2
Perfect score: 515
Sequence: 1 MTKTEKPFKLRSLFMPFH.....AQSAKQEVADSDAPASS 515

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

SPRMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 29 | 5.6 | 528 | 2 Q9SEV3 | Q9SEV3 chlamydia t |
| 2 | 23 | 2.5 | 501 | 16 Q92HP9 | Q92HP9 rickettsia |
| 3 | 11 | 2.1 | 105 | 10 Q945F3 | Q945F3 medicago sa |
| 4 | 11 | 2.1 | 588 | 8 Q8MFR8 | Q8MFR8 citrus hybr |
| 5 | 10 | 1.9 | 511 | 16 Q92HV4 | Q92HV4 rickettsia |
| 6 | 10 | 1.9 | 514 | 2 Q8KPN2 | Q8KPN2 holospira o |
| 7 | 10 | 1.9 | 644 | 10 Q9AS91 | Q9AS91 oryza sativ |
| 8 | 9 | 1.7 | 463 | 16 P95024 | P95024 mycobacteri |
| 9 | 9 | 1.7 | 463 | 16 Q8VUG6 | Q8VUG6 mycobacteri |
| 10 | 9 | 1.7 | 498 | 16 Q92JF6 | Q92JF6 rickettsia |
| 11 | 8 | 1.7 | 671 | 8 Q8MHS7 | Q8MHS7 galdieria s |
| 12 | 8 | 1.6 | 111 | 17 Q9YFB6 | Q9YFB6 aeropyrum p |
| 13 | 8 | 1.6 | 115 | 4 Q8NFU3 | Q8NFU3 homo sapien |
| 14 | 8 | 1.6 | 122 | 4 Q9BV22 | Q9BV22 homo sapien |
| 15 | 8 | 1.6 | 154 | 11 Q95362 | Q95362 mus musculu |
| 16 | 8 | 1.6 | 173 | 8 Q94992 | Q94992 elassoma ev |

| | | | | | |
|----|---|-----|------|-----------|---------------------|
| 17 | 8 | 1.6 | 194 | 16 Q24960 | Q24960 helicobacte |
| 18 | 8 | 1.6 | 197 | 16 Q9ZMT1 | Q9ZMT1 helicobacte |
| 19 | 8 | 1.6 | 217 | 4 Q96R14 | Q96R14 homo sapien |
| 20 | 8 | 1.6 | 238 | 4 Q8T063 | Q8T063 homo sapien |
| 21 | 8 | 1.6 | 245 | 6 Q95J13 | Q95J13 pan troglod |
| 22 | 8 | 1.6 | 247 | 15 Q85089 | Q85089 primate t-1 |
| 23 | 8 | 1.6 | 282 | 17 Q9YFUI | Q9YFUI aeropyrum p |
| 24 | 8 | 1.6 | 284 | 11 Q8K125 | Q8K125 mus musculu |
| 25 | 8 | 1.6 | 306 | 11 Q8VFP2 | Q8VFP2 mus musculu |
| 26 | 8 | 1.6 | 310 | 11 Q9WV14 | Q9WV14 mus musculu |
| 27 | 8 | 1.6 | 310 | 11 Q8VFB2 | Q8VFB2 mus musculu |
| 28 | 8 | 1.6 | 312 | 11 Q9WV13 | Q9WV13 mus musculu |
| 29 | 8 | 1.6 | 312 | 11 Q9WV11 | Q9WV11 mus musculu |
| 30 | 8 | 1.6 | 312 | 11 Q9R022 | Q9R022 mus musculu |
| 31 | 8 | 1.6 | 340 | 16 Q8F697 | Q8F697 leptospira |
| 32 | 8 | 1.6 | 342 | 4 Q9P042 | Q9P042 homo sapien |
| 33 | 8 | 1.6 | 356 | 4 Q9U021 | Q9U021 homo sapien |
| 34 | 8 | 1.6 | 356 | 4 Q96FY2 | Q96FY2 homo sapien |
| 35 | 8 | 1.6 | 357 | 4 Q60376 | Q60376 homo sapien |
| 36 | 8 | 1.6 | 360 | 10 Q9S1R6 | Q9S1R6 arabidopsis |
| 37 | 8 | 1.6 | 400 | 2 Q8GLH2 | Q8GLH2 photorhabdu |
| 38 | 8 | 1.6 | 480 | 16 Q8YV45 | Q8YV45 anabaena sp |
| 39 | 8 | 1.6 | 485 | 6 Q9GLN8 | Q9GLN8 pan troglod |
| 40 | 8 | 1.6 | 485 | 6 Q9GLP7 | Q9GLP7 pan troglod |
| 41 | 8 | 1.6 | 485 | 6 Q9GLP6 | Q9GLP6 gorilla gor |
| 42 | 8 | 1.6 | 491 | 15 Q8AGK3 | Q8AGK3 simian t-ly |
| 43 | 8 | 1.6 | 492 | 15 Q85091 | Q85091 primate t-1 |
| 44 | 8 | 1.6 | 510 | 16 Q8F082 | Q8F082 corynebacte |
| 45 | 8 | 1.6 | 520 | 11 Q921T2 | Q921T2 mus musculu |
| 46 | 8 | 1.6 | 545 | 10 Q8G0G3 | Q8G0G3 arabidopsis |
| 47 | 8 | 1.6 | 921 | 5 Q8MR77 | Q8MR77 drosophila |
| 48 | 8 | 1.6 | 967 | 2 Q8RK80 | Q8RK80 mycobacteri |
| 49 | 8 | 1.6 | 1013 | 3 Q9HF64 | Q9HF64 ashbya goss |
| 50 | 8 | 1.6 | 1360 | 5 Q9VD44 | Q9VD44 drosophila |
| 51 | 8 | 1.6 | 1772 | 16 Q9X877 | Q9X877 streptomyce |
| 52 | 8 | 1.6 | 3152 | 12 Q8U237 | Q8U237 leek yellow |
| 53 | 8 | 1.4 | 49 | 15 Q87390 | Q87390 chimpanzee |
| 54 | 8 | 1.4 | 58 | 16 Q8DS81 | Q8DS81 streptococ |
| 55 | 8 | 1.4 | 64 | 12 Q8QL33 | Q8QL33 sulfolobus |
| 56 | 8 | 1.4 | 82 | 13 Q9W6X0 | Q9W6X0 najia sputat |
| 57 | 8 | 1.4 | 83 | 13 Q57326 | Q57326 najia sputat |
| 58 | 8 | 1.4 | 83 | 13 Q9YGJ6 | Q9YGJ6 najia sputat |
| 59 | 8 | 1.4 | 83 | 13 Q9DE57 | Q9DE57 najia atra (|
| 60 | 8 | 1.4 | 83 | 13 Q9YGF5 | Q9YGF5 najia sputat |
| 61 | 8 | 1.4 | 83 | 13 Q9PTF0 | Q9PTF0 najia atra (|
| 62 | 8 | 1.4 | 83 | 13 Q57327 | Q57327 najia sputat |
| 63 | 8 | 1.4 | 84 | 16 Q99YR7 | Q99YR7 streptococ |
| 64 | 8 | 1.4 | 84 | 16 Q8P038 | Q8P038 streptococ |
| 65 | 8 | 1.4 | 84 | 16 Q8K6M5 | Q8K6M5 streptococ |
| 66 | 8 | 1.4 | 92 | 10 Q9ZTM9 | Q9ZTM9 petunia hyb |
| 67 | 8 | 1.4 | 101 | 8 Q99713 | Q99713 neomorphus |
| 68 | 8 | 1.4 | 102 | 2 Q52201 | Q52201 mycobacteri |
| 69 | 8 | 1.4 | 103 | 2 Q9ZNM6 | Q9ZNM6 onion yellow |
| 70 | 8 | 1.4 | 104 | 2 Q8VUR9 | Q8VUR9 aster yellow |
| 71 | 8 | 1.4 | 106 | 13 Q91988 | Q91988 xenopus lae |
| 72 | 8 | 1.4 | 106 | 13 Q96UF2 | Q96UF2 xenopus lae |
| 73 | 8 | 1.4 | 107 | 13 Q9PTT1 | Q9PTT1 gallus gall |
| 74 | 8 | 1.4 | 114 | 16 Q8PC77 | Q8PC77 xanthomonas |
| 75 | 8 | 1.4 | 116 | 16 Q86364 | Q86364 mycobacteri |
| 76 | 8 | 1.4 | 124 | 2 Q8GGE1 | Q8GGE1 streptomyce |
| 77 | 8 | 1.4 | 126 | 11 Q8BS28 | Q8BS28 mus musculu |
| 78 | 8 | 1.4 | 128 | 10 Q9M5M4 | Q9M5M4 euphorbia e |
| 79 | 8 | 1.4 | 136 | 5 Q8MNV2 | Q8MNV2 caenorhabdi |
| 80 | 8 | 1.4 | 137 | 16 Q8EYV1 | Q8EYV1 leptospira |
| 81 | 8 | 1.4 | 148 | 8 Q9MD16 | Q9MD16 scenesmus |
| 82 | 8 | 1.4 | 148 | 16 Q8EYV3 | Q8EYV3 leptospira |
| 83 | 8 | 1.4 | 152 | 17 Q58535 | Q58535 pyrococcus |
| 84 | 8 | 1.4 | 155 | 2 P72567 | P72567 streptococ |
| 85 | 8 | 1.4 | 155 | 16 Q9A1M6 | Q9A1M6 streptococ |
| 86 | 8 | 1.4 | 156 | 2 Q93QW8 | Q93QW8 pseudomonas |
| 87 | 8 | 1.4 | 157 | 12 Q83079 | Q83079 lychnis rin |
| 88 | 8 | 1.4 | 158 | 16 Q8ET47 | Q8ET47 oceanobacill |
| 89 | 8 | 1.4 | 159 | 8 Q8S1L8 | Q8S1L8 scaphidura |

90 1.4 161 5. 09VT62
91 1.4 166 2. 08X05
92 7 1.4 168 16. 08G00
93 7 1.4 173 8. 08WB67
94 7 1.4 173 8. 08LUL1
95 7 1.4 173 8. 0951J1
96 7 1.4 173 8. 0951I9
97 7 1.4 173 8. 08WB69
98 7 1.4 173 8. 094SD1
99 7 1.4 173 8. 094QGO
100 7 1.4 173 8. 08HM81

ALIGNMENTS

RESULT 1

0986V3 PRELIMINARY; PRT; 528 AA.
AC 0986V3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Nucleoside triphosphate transport protein 1.
GN NP1.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
CX NCBI_TaxID=813;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=99138740; PubMed=9973346;
RA Tjeden J., Winkler H.H., Schweeppe C., van der Laan M., Moehlmann T.,
Neubaus E.;
RT "Two nucleoside transport proteins in Chlamydia trachomatis : One for
net nucleoside triphosphate uptake and the other for the transport of
energy.";
RT J. Bacteriol. 181:1196-1202(1999).
RL EMBL; AJ010586; CAB39534.1; -
DR InterPro; IPR004667; ADP_ATP_car.
DR InterPro; IPR000531; TOMB_boxC.
DR Pfam; PF03219; T1C; 1.
DR TIGRFAMs; TIGR00769; AAA; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
SQ SEQUENCE 528 AA; 58099 MW; 8FECDD5FC08B0E3 CRC64;

Query Match 5.6%; Score 29; DB 2; Length 528;
Best local Similarity 100.0%; Pred. No. 7.1e-20;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 AYIPLDQKXKGAIDVVAARFGKSG 451
DB 421 AYIPLDQKXKGAIDVVAARFGKSG 449

RESULT 2

092HP9 PRELIMINARY; PRT; 501 AA.
AC 092HP9;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ADP_ATP carrier protein.
GN T1C3 OR RC0722.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiasee; Rickettsia.
CX NCBI_TaxID=781;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RA MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,

RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
DR EMBL; AE008630; AL03260.1; -
DR InterPro; IPR004667; ADP_ATP_car.
DR Pfam; PF03219; T1C; 1.
DR TIGRFAMs; TIGR00769; AAA; 1.
KW Complete proteome.
SQ SEQUENCE 501 AA; 57104 MW; 2EC589FAD865EA3B CRC64;

Query Match 2.5%; Score 13; DB 16; Length 501;
Best local Similarity 100.0%; Pred. No. 0.00066;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3

0945F3 PRELIMINARY; PRT; 105 AA.
AC 0945F3;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ADP/ATP translocase (Fragment).
OS Medicago sativa (Alfalfa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
CX NCBI_TaxID=3879;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Reagen S; TISSUE=Callus;
RA Winicov I.;
RT "Fragment of putative chloroplast ADP/ATP translocase from salt
tolerant callus.";
RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF16339; AL010405.1; -
DR InterPro; IPR004667; ADP_ATP_car.
DR Pfam; PF03219; T1C; 1.
FT NON_TER 1 1
FT NON_TER 105 105
SQ SEQUENCE 105 AA; 11212 MW; 1D0B51CC2C71B5A CRC64;

Query Match 2.1%; Score 11; DB 10; Length 105;
Best local Similarity 100.0%; Pred. No. 0.017;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 KVKGKAIDV 442
DB 51 KVKGKAIDV 61

RESULT 4

08MFR8 PRELIMINARY; PRT; 588 AA.
AC 08MFR8;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Plastidic ATP/ADP transporter (Fragment).
OS Citrus hybrid cultivar.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Sapindales; Rutaceae; Citrus.
CX NCBI_TaxID=171250;
RN (1)
RP SEQUENCE FROM N.A.
RA Li C., Weiss D., Goldschmidt E.E.;
RT "Citrus mRNA for plastidic ATP/ADP transporter.";

RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY098893; AAM29152.1; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; T1C; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KW Chloroplast.
 FT NON_TER
 SQ SEQUENCE 588 AA; 64735 MW; B4B6575BB36B37B CRC64;

Query Match 2.1%; Score 11; DB 8; Length 588;
 Best Local Similarity 100.0%; Pred. No. 0.076;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 VKGKAIDV 442
 |||||
 DB 483 VKGKAIDV 493

RESULT 5

092HV4 PRELIMINARY; PRT; 511 AA.
 AC 092HV4:
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE ADP_ATP carrier protein.
 GN T1C4 OR RC0666.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 CX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
 RL Science 293:2093-2098(2001).
 DR EMBL; AE008626; AAL03204.1; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; T1C; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KW Complete proteome.
 SQ SEQUENCE 511 AA; 58036 MW; E194D7E21591BE02 CRC64;

Query Match 1.9%; Score 10; DB 16; Length 511;
 Best Local Similarity 100.0%; Pred. No. 0.67;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 LFDSTKEMAY 424
 |||||
 DB 422 LFDSTKEMAY 431

RESULT 6

08KPN2 PRELIMINARY; PRT; 514 AA.
 AC 08KPN2:
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Non-mitochondrial nucleotide transport protein.
 GN NNT.
 OS Holospora obtusa.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Holosporaceae; Holospora.
 CX NCBI_TaxID=49893;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Linka N., Hurka H., Lang F.B., Burger G., Winkler H.H., Stamme C.,
 RA Urbany C., Seil I., Kuech J., Neuhau E.H.;

RT "Phylogenetic relationships of non-mitochondrial nucleotide transport
 RT proteins in bacteria and eukaryotes."
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY120885; AAM80566.1; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; T1C; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 SQ SEQUENCE 514 AA; 57531 MW; B6C3E24B9289E82F CRC64;

Query Match 1.9%; Score 10; DB 2; Length 514;
 Best Local Similarity 100.0%; Pred. No. 0.67;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 419 TREMAVYPLD 428
 |||||
 DB 423 TREMAVYPLD 432

RESULT 7

09AS91 PRELIMINARY; PRT; 644 AA.
 AC 09AS91:
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Putative plastidic ATP/ADP-transporter.
 GN P0707D10.36 OR P0038D11.15.
 OS Oryza sativa (Rice), and
 OS Oryza sativa (Japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eubacteriidae; Oryzae; Oryza.
 CX NCBI_TaxID=4530, 39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=O.sativa; STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0707D10.36";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=O.sativa (Japonica cultivar-group); STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0038D11.15";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP002910; BAB40979.1; -
 DR EMBL; AP003234; BAC05539.1; -
 DR Gramene; Q9AS91; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; T1C; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 SQ SEQUENCE 644 AA; 68960 MW; B3531863E48334AC CRC64;

Query Match 1.9%; Score 10; DB 10; Length 644;
 Best Local Similarity 100.0%; Pred. No. 0.82;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 433 VKGKAIDV 442
 |||||
 DB 526 VKGKAIDV 535

RESULT 8

P95024 PRELIMINARY; PRT; 463 AA.
 AC P95024:
 DT 01-MAY-1997 (Tremblrel. 03, Created)
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Hypothetical protein RV2529.
 GN RV2529 OR MTCY159.27C.

OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 Hornsby T., Jorgels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Ruter S., Seeger K., Skelton S., Squares S., Squares R.,
 Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544 (1998).
 DR EMBL: Z83663; CA806179.1; -
 DR TUBERCULISE; RV2529; -
 DR InterPro: IPR006166; ERCC4.
 DR Pfam: PF02732; ERCC4, 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 463 AA; 50192 MW; 5AC9AAB26394A97F CRC64;
 QY Query Match 1.7%; Score 9; DB 16; Length 463;
 DB Best Local Similarity 100.0%; Pred. No. 6.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 437 AADVVAAR 445
 DB 218 AADVVAAR 226
 RESULT 9
 OSVUG6 PRELIMINARY; PRT; 463 AA.
 ID OSVUG6
 AC OSVUG6
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein MT2604.
 GN MT2604
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 Delcher A., Utterback T., Weidman J., Khouri H., Gail J., Mikula A.,
 Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains";
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE007096; AAK46914.1; -
 DR TIGR: MT2604; -
 DR InterPro: IPR006166; ERCC4.
 DR Pfam: PF02732; ERCC4, 1.
 KW Hypothetical protein.
 SQ SEQUENCE 463 AA; 50201 MW; 1A7D6A019458E2B7 CRC64;
 QY Query Match 1.7%; Score 9; DB 16; Length 463;
 DB Best Local Similarity 100.0%; Pred. No. 6.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 437 AADVVAAR 445
 DB 218 AADVVAAR 226

RESULT 10
 ID Q92J16 PRELIMINARY; PRT; 498 AA.
 AC Q92J16
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE ADP/ATP carrier protein.
 GN TLC1 OR RC0081.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 Sausson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
 RL Science 293:2093-2098 (2001).
 DR EMBL: AE008575; AL02619.1; -
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam: PF03219; TLC; 1.
 DR TIGRFAMs: TIGR00769; AAA; 1.
 KW Complete proteome.
 SQ SEQUENCE 498 AA; 56661 MW; 97B5EB9C5BC56B CRC64;
 QY Query Match 1.7%; Score 9; DB 16; Length 498;
 DB Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 246 NKNVLTDP 254
 DB 241 NKNVLTDP 249
 RESULT 11
 OSVH67 PRELIMINARY; PRT; 671 AA.
 ID OSVH67
 AC OSVH67
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Plastidic ATP/ADP transporter (Fragment).
 GN ATP1.
 OS Galdieria sulphuraria.
 OC Chloroplast.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
 OC Galdieria.
 OX NCBI_TaxID=130081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stehme C., Neuhäus E.H.;
 RT "Analysis of the plastidic ATP/ADP transporter from the red algae
 R. Galdieria sulphuraria";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ251356; CAC80882.1; -
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam: PF03219; TLC; 1.
 DR TIGRFAMs: TIGR00769; AAA; 1.
 KW Chloroplast.
 FT NON TER
 SQ SEQUENCE 671 AA; 73829 MW; 2D25A868F2CFFD7 CRC64;
 QY Query Match 1.7%; Score 9; DB 8; Length 671;
 DB Best Local Similarity 100.0%; Pred. No. 8.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 390 GTPLMLAV 398

Db 534 GTPPLMLAV 542

|||||

RESULT 12

ID Q9YFB6 PRELIMINARY; PRT; 111 AA.
AC Q9YFB6
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein APE0325.
GN APE0325.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcaceae;
OC Desulfurococcaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=9310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Hatakeyama Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankal A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Ogunchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000059; BAA79280.1; -
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 111 AA; 10742 MW; CP9180F4DF1A27B8 CRC64;

Query Match 1.6%; Score 8; DB 17; Length 111;
Best Local Similarity 100.0%; Pred. No. 18;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 LAVVNGAI 403

Db 3 LAVVNGAI 10

RESULT 13

ID Q8NFU3 PRELIMINARY; PRT; 115 AA.
AC Q8NFU3
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE KAT protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wenzel K., Hohe M.R., Bauer D.,
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF394442; AAM69839.1; -
DR InterPro; IPR001763; Rhodanese-like.
DR Pfam; PF00581; Rhodanese; 1.
DR SMART; SM00450; RHOD; 1.
DR PROSITE; PS50206; RHODANES; 1.
SQ SEQUENCE 115 AA; 12530 MW; 01963F5B8186B3A2 CRC64;

Query Match 1.6%; Score 8; DB 4; Length 115;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 SLASGRA 202

Db 14 SLASGRA 21

RESULT 14

ID Q9BV22 PRELIMINARY; PRT; 122 AA.
AC Q9BV22
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Lymph;
RA Strausberg R.,
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC001699; AAH01699.1; -
DR InterPro; IPR001763; Rhodanese-like.
DR Pfam; PF00581; Rhodanese; 1.
DR SMART; SM00450; RHOD; 1.
DR PROSITE; PS50206; RHODANES; 1.
KM Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 122 AA; 13187 MW; 0C63813B165ADE06 CRC64;

Query Match 1.6%; Score 8; DB 4; Length 122;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 SLASGRA 202

Db 27 SLASGRA 34

RESULT 15

ID O35362 PRELIMINARY; PRT; 154 AA.
AC O35362
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Casr-rs1 (Fragment).
GN CASR-RS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98008915; PubMed=9344650;
RA Hinson T.K., Damodaran T.V., Chen J., Zhang X., Qumaiyeh M.B.,
RA Seldin M.F., Quarles L.D.;
RT "Identification of putative transmembrane receptor sequences
RT homologous to the calcium-sensing G-protein-coupled receptor."
RL Genomics 45:279-289(1997).
DR EMBL; AF022250; AAB83974.1; -
DR InterPro; IPR001091; GPCR_MGR.
DR InterPro; IPR000337; GPCR_MGR.
DR InterPro; IPR004073; Vmron_receptor2.
DR Pfam; PF00003; 7tm_3; 1.
DR PRINTS; PR00248; GPCR_MGR.
DR PRINTS; PR01535; VOMERONASL2R.
DR PROSITE; PS50259; G_PROTEIN_RECP_F3_4; 1.
DR PROSITE; PS00093; N4_MTASE; 1.
FT NON_TER 1
SQ SEQUENCE 154 AA; 16875 MW; 325F445080A1CB81 CRC64;

Query Match 1.6%; Score 8; DB 11; Length 154;
Best Local Similarity 100.0%; Pred. No. 24;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 286 YIILTL 293
 Db 7 YIILTL 14

RESULT 16

Q94S92 PRELIMINARY; PRT; 173 AA.

AC Q94S92; 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

GN MD6. NADH dehydrogenase subunit 6.

OS Elasmobranch evergladesi (Everglades pygmy sunfish).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorphi; Acanthopterygii; Percormorpha; Elasmobranchia; Elasmobranchia.

NCBI_TaxID=71674;

Qy 286 YIILTL 293

Db 150 YIILTL 157

Qy 287 YIILTL 294

Db 150 YIILTL 157

Qy 287 YIILTL 294

Db 150 YIILTL 157

Qy 287 YIILTL 294

Db 150 YIILTL 157

Qy 287 YIILTL 294

Db 150 YIILTL 157

Qy 287 YIILTL 294

Db 150 YIILTL 157

Qy 287 YIILTL 294

Db 150 YIILTL 157

Qy 287 YIILTL 294

Db 150 YIILTL 157

Qy 287 YIILTL 294

Db 150 YIILTL 157

Qy 287 YIILTL 294

Db 150 YIILTL 157

Qy 287 YIILTL 294

Db 150 YIILTL 157

Qy 287 YIILTL 294

RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 RT pylori.";
 RL Nature 388:539-547(1997).
 DR EMBL; AE000536; AAD07223.1; -.
 DR TIGR; HP0149; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 194 AA; 22239 MW; B06EA12F553BD809 CRC64;

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

Qy 286 YIILTL 293

Db 175 YIILTL 182

```

RT "DEFOG: A Practical Scheme for Deciphering Families of Genes.";
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF399631; AAK5116.1; -.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHDODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECPE_FL_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECPE_FL_2; 1.
KW Receptor.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 217 AA; 24112 MW; D0855DEDE6F09384 CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 4; Length 217;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 LFYAVGTP 101
Db 209 LFYAVGTP 216

RESULT 20
Q8IU63 PRELIMINARY; PRT; 238 AA.
ID Q8IU63
AC Q8IU63
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Olfactory receptor.
GN 6M1-16.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Voiz A., Ehlers A., Younger R., Forbes S., Trowsdale J., Beck S.,
RA Ziegler A.;
RT "Complex transcriptional control of MHC-linked olfactory receptor
RT genes includes long distance and extensive alternative splicing; exon
RT sharing and premature polyadenylation.";
RL Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ459845; CAD31037.1; -.
DR EMBL; AJ459846; CAD31038.1; -.
KW Receptor.
SQ SEQUENCE 238 AA; 26550 MW; B9BB34DBBF6328EF CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 4; Length 238;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 LFYAVGTP 101
Db 196 LFYAVGTP 203

RESULT 21
Q95J13 PRELIMINARY; PRT; 245 AA.
ID Q95J13
AC Q95J13
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Angiotensinogen (Fragment).
GN REN.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
OC NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=504, and 505;
RA Sarta Y.;
RT "Comparison of DNA and protein polymorphisms between humans and
RT chimpanzees.";
RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
CC -1-SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL; AB062027; BAB55856.1; -.
DR EMBL; AB062028; BAB55857.1; -.
DR InterPro; IPR000227; Angiotensn.
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin; 1.
DR PRINTS; PR00654; ANGIOTENNGN.
KW Protease inhibitor; Serine protease inhibitor; Serpin.
FT NON_TER 245
SQ SEQUENCE 245 AA; 26317 MW; E0092390B9803E0B CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 6; Length 245;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 ADRLOAIL 133
Db 155 ADRLOAIL 162

RESULT 22
Q85089 PRELIMINARY; PRT; 247 AA.
ID Q85089
AC Q85089
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Env protein (Fragment).
GN ENV.
OS Primate T-lymphotropic virus 3.
OC Viruses; Retroid viruses; Retroviridae; Deltaretrovirus.
OC NCBI_TaxID=194443;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=STLV-PH 969;
RX MEDLINE=94195839; PubMed=7908445;
RA Goubau P., Van Brussel M., Vandamme A., Liu H.F., Desmyter J.;
RT "A primate T-lymphotropic virus, PTLV-L, different from human T-
RT lymphotropic viruses type I and II, in a wild-caught baboon (Papio
RT hamadryas).";
RL Proc. Natl. Acad. Sci. U.S.A. 91:2848-2852(1994).
DR EMBL; Z29673; CA82771.1; -.
DR HSSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON_TER 1
SQ SEQUENCE 247 AA; 27207 MW; B8EDA1B7B951FE54 CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 15; Length 247;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 480 IAIWLVA 487
Db 73 IAIWLVA 80

RESULT 23
Q9YFUL PRELIMINARY; PRT; 282 AA.
ID Q9YFUL
AC Q9YFUL
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE 282AA long hypothetical bacteriochlorophyll synthase.
GN APE0159.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcates;

```

```

OC Desulfurococcaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,
RA Jinno K., Takahashi M., Sekine M., Baba S.-I., Anai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kusuda N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000058; BAA79070.1; -.
DR InterPro; IPR000812; TFIIB_euk.
DR InterPro; IPR000537; UblA.
DR Pfam; PF01040; UblA; 1.
DR PROSITE; PS00782; TFIIB; 1.
KW Complete proteome.
SQ SEQUENCE 282 AA; 29648 MW; D39C1EAD0B9E600 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 17; Length 282;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILLTLTLV 294
DB 234 ILLTLTLV 241

RESULT 24
Q8K125 PRELIMINARY; PRT; 294 AA.
AC Q8K125;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
DS Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC034887; AAH34887.1; -.
DR InterPro; IPR003599; IG_1-like.
DR InterPro; IPR007110; IG_1-like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00409; IG; 1.
DR SMART; SM00407; IGc1; 1.
DR PROSITE; PS50835; IG_LIKE; 2.
KW Hypothetical protein.
SQ SEQUENCE 294 AA; 32368 MW; 6DBBEB36659735F2 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 11; Length 294;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGLVIMA 240
DB 280 VSGLVIMA 287

RESULT 25
Q8VF92 PRELIMINARY; PRT; 306 AA.
ID Q8VF92

```

```

AC Q8VF92;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Olfactory receptor MOR256-29.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang X., Firestein S.J.;
RT "The olfactory receptor gene superfamily of the mouse.";
RL Nat. Neurosci. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Adams M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY073642; AAL61305.1; -.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODOPS.
DR PROSITE; PS00237; G_PROTEIN_RECIP_F1_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECIP_F1_2; 1.
KW Receptor.
SQ SEQUENCE 306 AA; 34054 MW; 361F56047B31E9CA CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 11; Length 306;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
DB 274 LFYAVGTP 281

RESULT 26
Q9WV14 PRELIMINARY; PRT; 310 AA.
AC Q9WV14;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 573K1.2 (Mm17M1-3 (Novel 7 transmembrane receptor (Rhodopsin family)
DE (Olfactory receptor like) protein)).
DE GN GABRI OR 573K1.2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Younger R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078630; CAB44994.1; -.
DR MGD; MGI:1860139; Gabbr1.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODOPS.
DR PROSITE; PS00237; G_PROTEIN_RECIP_F1_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECIP_F1_2; 1.
KW Receptor; Transmembrane.
SQ SEQUENCE 310 AA; 34540 MW; 8C069BF46E35DEAD CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 11; Length 310;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
DB 274 LFYAVGTP 281

```


RESULT 27

Q9VFB2 PRELIMINARY; PRT; 310 AA.
AC Q9VFB2; 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Olfactory receptor MOR256-21.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
GN NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang X., Firestein S.J.;
RT "The olfactory receptor gene superfamily of the mouse."
RL Nat. Neurosci. 0:0-0(2002).
RN [12]
RP SEQUENCE FROM N.A.
RA Adams M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY073621; AAL61284.1; -.
DR InterPro; IPR000276; GPCR_Rhodpn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHOOPS.
DR PROSITE; PS00237; G_PROTEIN_RECP_F1_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECP_F1_2; 1.
KM Receptor.
SQ SEQUENCE 310 AA; 34641 MW; 0C857283AD7A66 CRC64;

Query Match 1.6%; Score 8; DB 11; Length 310;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
Db 274 LFYAVGTP 281

RESULT 28

Q9WV13 PRELIMINARY; PRT; 312 AA.
AC Q9WV13; 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 573K1.4 (Mm17M1-1 (Novel 7 transmembrane receptor (Rhodopsin family)
GN GABBR1 OR 573K1.4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
GN NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Younger R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078630; CAB44996.1; -.
DR MGD; MGI:1860139; Gabbri1.
DR InterPro; IPR000276; GPCR_Rhodpn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHOOPS.
DR PROSITE; PS00237; G_PROTEIN_RECP_F1_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECP_F1_2; 1.
KM Receptor; Transmembrane.
SQ SEQUENCE 312 AA; 34988 MW; 630498DA34FCD1D0 CRC64;

Query Match 1.6%; Score 8; DB 11; Length 312;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
Db 274 LFYAVGTP 281

Db 274 LFYAVGTP 281

RESULT 29

Q9WV11 PRELIMINARY; PRT; 312 AA.
AC Q9WV11; 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 573K1.8 (Mm17M1-2 (Novel 7 transmembrane receptor (rhodopsin family)
GN GABBR1 OR 573K1.8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
GN NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Younger R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078630; CAB45000.1; -.
DR MGD; MGI:1860139; Gabbri1.
DR InterPro; IPR000276; GPCR_Rhodpn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHOOPS.
DR PROSITE; PS00237; G_PROTEIN_RECP_F1_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECP_F1_2; 1.
KM Receptor; Transmembrane.
SQ SEQUENCE 312 AA; 34786 MW; DAA7507C77710AD8 CRC64;

Query Match 1.6%; Score 8; DB 11; Length 312;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
Db 274 LFYAVGTP 281

RESULT 30

Q9R0Z2 PRELIMINARY; PRT; 312 AA.
AC Q9R0Z2; 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 573K1.3 (Mm17M1-4 (Novel 7 transmembrane receptor (rhodopsin family)
GN GABBR1 OR 573K1.3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
GN NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Younger R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078630; CAB44996.1; -.
DR MGD; MGI:1860139; Gabbri1.
DR InterPro; IPR000276; GPCR_Rhodpn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHOOPS.
DR PROSITE; PS00237; G_PROTEIN_RECP_F1_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECP_F1_2; 1.
KM Receptor; Transmembrane.
SQ SEQUENCE 312 AA; 34988 MW; 630498DA34FCD1D0 CRC64;

Query Match 1.6%; Score 8; DB 11; Length 312;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_FL_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECEP_FL_2; 1.
DR RECEPTOR; Transmembrane.
SQ SEQUENCE 312 AA; 3499 MW; 83374B1A4C6F912D CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 11; Length 312;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFVAVGTP 101
Db 274 LFVAVGTP 281

RESULT 31
Q8F697 PRELIMINARY; PRT; 340 AA.
AC Q8F697;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN IAI412.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
CX NCBI_TaxID=173;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE011321; AAA8611.1.-.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 340 AA; 38370 MW; 1A0520FA1B5DC0F CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 16; Length 340;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 LGLVAIL 144
Db 318 LGLVAIL 325

RESULT 32
Q9P042 PRELIMINARY; PRT; 342 AA.
AC Q9P042;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE HSC108.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RL MEDLINE=20499367; PubMed=11042152;
RA Zhang Q.H., Ye M., Wu X.Y., Ren S.X., Zhao M., Zhao C.J., Fu G.,
RA Shen Y., Fan H.Y., Lu G., Zhong M., Xu X.R., Han Z.G., Zhang J.W.,
RA Tao J., Huang Q.H., Zhou J., Hu G.X., Gu J., Chen S.D., Chen Z.;
RT "Cloning and functional analysis of cDNAs with open reading frames for
RT 300 previously undefined genes expressed in CD34+ hematopoietic
RT stem/progenitor cells.";
RL Genome Res. 10:1546-1560(2000).
DR EMBL; AF161458; AAF29073.1; -.
DR InterPro; IPR001107; Band_7.
DR InterPro; IPR001972; Stomatin.
DR Pfam; PF01145; Band_7; 1.
DR PRINTS; PR00721; STOMATIN.

```

```

DR SMART; SM00244; PHB; 1.
SQ SEQUENCE 342 AA; 37145 MW; 7958C0B3BDE53E5 CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 4; Length 342;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 SLASGRA 202
Db 4 SLASGRA 11

RESULT 33
Q9UZ1 PRELIMINARY; PRT; 356 AA.
AC Q9UZ1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Membrane associated protein SLP-2 (Stomatin-like protein 2) (Stomatin-
DE like 2) (Hypothetical protein FLJ14499).
GN HUSLP2 OR SLP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Heart muscle;
RA Wang Y., Morrow J.S.;
RT "Identification and Characterization of Human SLP-2, a Novel Homologue
RT of Stomatin (band 7.2b) Present in Erythrocytes and Other Tissues.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Owczarek C.M., Treutlein H.R., Portbury K.J., Gulluyan L.M., Kola I.,
RA Hertzog P.J.;
RT "A novel member of the stomatin/EPB72/MC-2 family, stomatin-like
RT protein 2 (SLP2), is ubiquitously expressed and localizes to HSA
RL chromosome 9p13.1.";
RN (3)
RP SEQUENCE FROM N.A.
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
RA Wagaetsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Iehi S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Niromiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN (4)
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas, Lung, and Skin;
RA Straubeberg R.;
RL EMBL; AF190167; AAF09142.1; -.
DR EMBL; AF190167; AAF09142.1; -.
DR EMBL; AF282596; AAF91466.1; -.
DR EMBL; BC003025; AAH03025.1; -.
DR EMBL; BC002442; AAH02442.1; -.
DR EMBL; AK027405; BAB5091.1; -.
DR EMBL; BC014990; AAH14990.1; -.
DR InterPro; IPR001107; Band_7.
DR InterPro; IPR001972; Stomatin.
DR Pfam; PF01145; Band_7; 1.
DR PRINTS; PR00721; STOMATIN.
DR SMART; SM00244; PHB; 1.
KW Hypothetical protein.
SQ SEQUENCE 356 AA; 38534 MW; 672331B57C82654E CRC64;

Query Match
1.6%; Score 8; DB 4; Length 356;

```

Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 195 SLASGRA 202
Db 17 SLASGRA 24

RESULT 34

Q96FY2 PRELIMINARY; PRT; 356 AA.
AC Q96FY2; 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Stomatin-like 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010152; AAH10152.1; -
DR InterPro; IPR001107; Band_7;
DR Pfam; PF01145; Band_7; 1.
DR PRINTS; PR00721; STOMATIN.
DR SMART; SM00244; PHB; 1.
SQ SEQUENCE 356 AA; 38518 MW; 6FC0BF899E1919A7 CRC64;

Query Match 1.6%; Score 8; DB 4; Length 356;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 195 SLASGRA 202
Db 17 SLASGRA 24

RESULT 35

060376 PRELIMINARY; PRT; 357 AA.
AC 060376; 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fibroblast or foreskin.
RA Lamerzin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stiliwgen S.,
RA Phan H., Velasco N., Garmes J., Dangnan L., Poundstone P.,
RA Christensen M., Georgescu A., Avila J., Liu S., Altix C., Andreise T.,
RA Frankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
RA Montgomey M., Ow D., Nolan M., Triong S., Kobayashi A., Olsen A.O.,
RA Carrano A.V.;
RT "Sequence analysis of a human P1 clone containing the XRC9 DNA repair
gene";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004472; AAC07983.1; -
DR Genew; HGN14559; STOML2.
DR InterPro; IPR001107; Band_7.
DR Pfam; PF01145; Band_7; 1.

DR PRINTS; PR00721; STOMATIN.
DR SMART; SM00244; PHB; 1.
SQ SEQUENCE 357 AA; 38749 MW; 512632B83028135A CRC64;

Query Match 1.6%; Score 8; DB 4; Length 357;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 195 SLASGRA 202
Db 4 SLASGRA 11

RESULT 36

Q9SIR6 PRELIMINARY; PRT; 360 AA.
AC Q9SIR6; 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE At2g25270 protein.
GN At2g25270.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Beell C.R., Ketchum K.A., Lee J.-J., Rouning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Umeyam L., Tallon L.J., Gall J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana";
RL Nature 402:761-768 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007070; AAD23664.1; -
SQ SEQUENCE 360 AA; 39992 MW; EAPD5D2A9E57870F CRC64;

Query Match 1.6%; Score 8; DB 10; Length 360;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 162 VMLSLMF 169
Db 325 VMLSLMF 332

RESULT 37

08GLH2 PRELIMINARY; PRT; 400 AA.
AC 08GLH2; 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Acetate kinase (EC 2.7.2.1).
GN ACK.
OS Photobacterium temperata.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photobacterium.
OX NCBI_Taxid=171441;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=K122;
RA Joyce S.A., Clarke D.J.;
RT "A hexa homolog from Phototrichobius regulates pathogenicity, symbiosis
RT and phenotypic variation.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY137386; AAN08359.1; -.
KW Kinase; Transferase.
SQ SEQUENCE 400 AA; 43388 MW; AD4D6555618F1DF CRC64;

Query Match 1.6%; Score 8; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 443 AARFGKSG 450
Db 363 AARFGKSG 370

RESULT 38
Q8YV45 PRELIMINARY; PRT; 480 AA.
AC Q8YV45;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein A111007.
GN A111007.
OS Arabidopsis sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneo T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003584; BAB72964.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 480 AA; 54939 MW; 22908DF99BB417A CRC64;

Query Match 1.6%; Score 8; DB 16; Length 480;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 268 LLTLTLLV 295
Db 429 LLTLTLLV 436

RESULT 39
Q9GLN8 PRELIMINARY; PRT; 485 AA.
AC Q9GLN8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Angiotensinogen.
GN AGT.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20469400; PubMed=11013071;
RA Dufour C., Casane D., Denton D., Wickings J., Corvol P.,
RA Jeunemaitre X.;
RT "Human-Chimpanzee DNA sequence variation in the four major genes of

```

```

RT the renin angiotensin system.";
RL Genomics 69:14-26(2000).
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL; AF193461; AAG30306.1; -.
DR EMBL; AF193458; AAG30306.1; JOINED.
DR EMBL; AF193459; AAG30306.1; JOINED.
DR EMBL; AF193460; AAG30306.1; JOINED.
DR InterPro; IPR000227; Angiotensn.
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin.1.
DR PRINTS; PR00654; ANGIOTENSGN.
DR SMART; SM00093; SERPIN.1.
DR PROSITE; PS00284; SERPIN.1.
KW Protease inhibitor; Serine protease inhibitor; Serpin.
SQ SEQUENCE 485 AA; 53110 MW; C14C67B49A53F05F CRC64;

Query Match 1.6%; Score 8; DB 6; Length 485;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 ADRLOAIL 133
Db 155 ADRLOAIL 162

RESULT 40
Q9GLP7 PRELIMINARY; PRT; 485 AA.
AC Q9GLP7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Angiotensinogen.
GN AGT.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX Shattuck-Eidens D., McGrath M., Stone S.;
RT "Germline mutations in the angiotensinogen gene cause predisposition
RT to type 1 diabetes mellitus.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL; AF188467; AAG29056.1; -.
DR InterPro; IPR000227; Angiotensn.
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin.1.
DR PRINTS; PR00654; ANGIOTENSGN.
DR SMART; SM00093; SERPIN.1.
DR PROSITE; PS00284; SERPIN.1.
KW Protease inhibitor; Serine protease inhibitor; Serpin.
SQ SEQUENCE 485 AA; 53140 MW; 49EFB54AF31F8ADC CRC64;

Search completed: November 25, 2003, 10:07:00
Job time : 70 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: November 25, 2003, 10:05:06 ; Search time 23 Seconds
(without alignments)
947.395 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 515
Sequence: 1 MKTEKPKPKLRSLFMPH.....AGSLKEQVAGDSAPASS 515

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/aa/5A COMB.pep:*
2: /cgn2_6/prodata/1/aa/5B COMB.pep:*
3: /cgn2_6/prodata/1/aa/6A COMB.pep:*
4: /cgn2_6/prodata/1/aa/6B COMB.pep:*
5: /cgn2_6/prodata/1/aa/PCTUS COMB.pep:*
6: /cgn2_6/prodata/1/aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------------------|--------------------|
| 1 | 313 | 60.8 | 524 | US-09-198-452A-369 | Sequence 369, App |
| 2 | 8 | 1.6 | 111 | US-09-328-352-5992 | Sequence 5992, Ap |
| 3 | 8 | 1.6 | 137 | US-08-974-546-3 | Sequence 3, Appl |
| 4 | 8 | 1.6 | 172 | US-09-252-991A-23710 | Sequence 23710, A |
| 5 | 8 | 1.6 | 293 | US-09-509-347-7 | Patent No. 5189147 |
| 6 | 8 | 1.6 | 303 | US-09-509-347-7 | Sequence 7, Appl |
| 7 | 8 | 1.6 | 485 | US-09-384-212-2 | Sequence 2, Appl |
| 8 | 8 | 1.6 | 1013 | US-09-415-522-8 | Sequence 8, Appl |
| 9 | 8 | 1.6 | 1964 | US-09-467-997-1 | Sequence 1, Appl |
| 10 | 8 | 1.6 | 16 | US-07-894-063A-1 | Sequence 1, Appl |
| 11 | 8 | 1.6 | 16 | US-07-894-063A-2 | Sequence 2, Appl |
| 12 | 8 | 1.6 | 77 | US-09-393-634-11 | Sequence 11, Appl |
| 13 | 8 | 1.6 | 110 | US-09-732-210-1020 | Sequence 1020, Ap |
| 14 | 8 | 1.6 | 117 | US-08-444-818-44 | Sequence 44, Appl |
| 15 | 8 | 1.6 | 132 | US-08-444-818-52 | Sequence 52, Appl |
| 16 | 8 | 1.6 | 159 | US-09-252-991A-29736 | Sequence 29736, A |
| 17 | 8 | 1.6 | 193 | US-08-308-821A-6 | Sequence 6, Appl |
| 18 | 8 | 1.6 | 193 | US-08-831-627-6 | Sequence 6, Appl |
| 19 | 8 | 1.6 | 193 | US-08-765-381-15 | Sequence 15, Appl |
| 20 | 8 | 1.6 | 197 | US-08-604-372A-32 | Sequence 32, Appl |
| 21 | 8 | 1.6 | 211 | PCT-US94-04174-18 | Sequence 18, Appl |
| 22 | 8 | 1.6 | 216 | US-09-252-991A-21062 | Sequence 21062, A |
| 23 | 8 | 1.6 | 228 | US-09-328-352-4385 | Sequence 4385, Ap |
| 24 | 8 | 1.6 | 225 | US-09-107-532A-7302 | Sequence 7302, Ap |
| 25 | 8 | 1.6 | 253 | US-09-333-599-2 | Sequence 2, Appl |
| 26 | 8 | 1.6 | 253 | US-09-499-781-2 | Sequence 2, Appl |
| 27 | 8 | 1.6 | 259 | US-09-252-991A-19115 | Sequence 19115, A |

| | | | | | |
|-----|---|-----|------|----------------------|-------------------|
| 28 | 7 | 1.4 | 279 | US-09-328-352-7144 | Sequence 7144, Ap |
| 29 | 7 | 1.4 | 293 | US-09-414-276-3 | Sequence 3, Appl |
| 30 | 7 | 1.4 | 317 | US-09-134-001C-4537 | Sequence 4537, Ap |
| 31 | 7 | 1.4 | 335 | US-09-414-276-6 | Sequence 6, Appl |
| 32 | 7 | 1.4 | 357 | US-09-107-532A-5132 | Sequence 5132, Ap |
| 33 | 7 | 1.4 | 362 | US-09-252-991A-18068 | Sequence 18068, A |
| 34 | 7 | 1.4 | 375 | US-09-622-439-2 | Sequence 2, Appl |
| 35 | 7 | 1.4 | 414 | US-09-107-532A-6371 | Sequence 6371, Ap |
| 36 | 7 | 1.4 | 423 | US-09-328-352-6097 | Sequence 6097, Ap |
| 37 | 7 | 1.4 | 423 | US-09-328-352-7964 | Sequence 7964, Ap |
| 38 | 7 | 1.4 | 452 | US-09-252-991A-20578 | Sequence 20578, A |
| 39 | 7 | 1.4 | 504 | US-07-853-985A-6 | Sequence 6, Appl |
| 40 | 7 | 1.4 | 504 | US-07-681-703B-6 | Sequence 6, Appl |
| 41 | 7 | 1.4 | 504 | US-08-184-236-6 | Sequence 6, Appl |
| 42 | 7 | 1.4 | 504 | US-08-407-410B-6 | Sequence 6, Appl |
| 43 | 7 | 1.4 | 504 | US-08-485-500-6 | Sequence 6, Appl |
| 44 | 7 | 1.4 | 504 | PCT-US91-02370-6 | Sequence 6, Appl |
| 45 | 7 | 1.4 | 504 | PCT-US94-04174-6 | Sequence 6, Appl |
| 46 | 7 | 1.4 | 532 | US-09-252-991A-27288 | Sequence 27288, A |
| 47 | 7 | 1.4 | 541 | US-08-867-611-8 | Sequence 8, Appl |
| 48 | 7 | 1.4 | 541 | PCT-US92-06965A-13 | Sequence 13, Appl |
| 49 | 7 | 1.4 | 574 | US-08-309-512-9 | Sequence 9, Appl |
| 50 | 7 | 1.4 | 574 | PCT-US92-08766A-9 | Sequence 9, Appl |
| 51 | 7 | 1.4 | 579 | US-09-173-151A-2 | Sequence 2, Appl |
| 52 | 7 | 1.4 | 590 | US-09-208-140-4 | Sequence 4, Appl |
| 53 | 7 | 1.4 | 593 | US-09-208-140-2 | Sequence 2, Appl |
| 54 | 7 | 1.4 | 593 | US-08-722-806A-2 | Sequence 2, Appl |
| 55 | 7 | 1.4 | 593 | US-09-337-028-2 | Sequence 2, Appl |
| 56 | 7 | 1.4 | 593 | US-09-597-877-2 | Sequence 2, Appl |
| 57 | 7 | 1.4 | 604 | US-09-345-473B-17 | Sequence 17, Appl |
| 58 | 7 | 1.4 | 646 | US-09-614-259-33 | Sequence 33, Appl |
| 59 | 7 | 1.4 | 785 | US-09-252-991A-18587 | Sequence 18587, A |
| 60 | 7 | 1.4 | 785 | US-09-252-991A-32952 | Sequence 32952, A |
| 61 | 7 | 1.4 | 798 | US-08-867-611-36 | Sequence 36, Appl |
| 62 | 7 | 1.4 | 919 | US-09-437-054A-17 | Sequence 17, Appl |
| 63 | 7 | 1.4 | 1011 | US-08-850-328-5 | Sequence 5, Appl |
| 64 | 7 | 1.4 | 1786 | US-08-444-818-54 | Sequence 54, Appl |
| 65 | 7 | 1.4 | 2261 | US-08-444-818-66 | Sequence 66, Appl |
| 66 | 7 | 1.4 | 2436 | US-08-444-818-75 | Sequence 75, Appl |
| 67 | 7 | 1.4 | 2722 | US-08-444-818-89 | Sequence 89, Appl |
| 68 | 7 | 1.4 | 2955 | US-08-442-805A-3 | Sequence 3, Appl |
| 69 | 7 | 1.4 | 2955 | US-08-442-805A-3 | Sequence 3, Appl |
| 70 | 7 | 1.4 | 2955 | US-08-442-805A-3 | Sequence 3, Appl |
| 71 | 7 | 1.4 | 2955 | US-08-444-818-124 | Sequence 124, App |
| 72 | 7 | 1.4 | 2955 | US-08-249-843-3 | Sequence 3, Appl |
| 73 | 7 | 1.4 | 2995 | US-08-444-818-138 | Sequence 138, App |
| 74 | 7 | 1.4 | 3011 | US-08-188-281B-1 | Sequence 1, Appl |
| 75 | 7 | 1.4 | 3011 | US-08-453-552-1 | Sequence 1, Appl |
| 76 | 7 | 1.4 | 3011 | US-08-440-103-36 | Sequence 36, Appl |
| 77 | 7 | 1.4 | 3011 | US-08-440-103-36 | Sequence 36, Appl |
| 78 | 7 | 1.4 | 3011 | US-08-440-103-36 | Sequence 36, Appl |
| 79 | 7 | 1.4 | 3011 | US-07-910-760-10 | Sequence 10, Appl |
| 80 | 7 | 1.4 | 3011 | US-08-440-519-10 | Sequence 10, Appl |
| 81 | 7 | 1.4 | 3011 | US-08-231-368-36 | Sequence 36, Appl |
| 82 | 7 | 1.4 | 3011 | US-08-440-210-36 | Sequence 36, Appl |
| 83 | 7 | 1.4 | 3011 | US-08-710-637-1 | Sequence 1, Appl |
| 84 | 7 | 1.4 | 3011 | US-08-833-678A-6 | Sequence 6, Appl |
| 85 | 7 | 1.4 | 3011 | US-08-811-566-20 | Sequence 20, Appl |
| 86 | 7 | 1.4 | 3011 | US-08-444-818-177 | Sequence 177, App |
| 87 | 7 | 1.4 | 3011 | US-09-014-416-1 | Sequence 1, Appl |
| 88 | 7 | 1.4 | 3011 | US-09-014-416-5 | Sequence 5, Appl |
| 89 | 7 | 1.4 | 3011 | US-08-529-169A-6 | Sequence 6, Appl |
| 90 | 7 | 1.4 | 3011 | US-09-388-874-2 | Sequence 2, Appl |
| 91 | 7 | 1.4 | 3011 | US-09-046-604-36 | Sequence 36, Appl |
| 92 | 7 | 1.4 | 3011 | US-08-440-549-10 | Sequence 10, Appl |
| 93 | 7 | 1.4 | 3011 | US-08-850-328-1 | Sequence 1, Appl |
| 94 | 7 | 1.4 | 3011 | US-09-034-756-20 | Sequence 20, Appl |
| 95 | 7 | 1.4 | 3011 | US-09-483-799-6 | Sequence 6, Appl |
| 96 | 7 | 1.4 | 3011 | US-09-916-359-2 | Sequence 2, Appl |
| 97 | 7 | 1.4 | 3011 | PCT-US91-02225-10 | Sequence 10, Appl |
| 98 | 7 | 1.4 | 3011 | PCT-US93-00907-1 | Sequence 1, Appl |
| 99 | 7 | 1.4 | 3011 | PCT-US94-07280-1 | Sequence 1, Appl |
| 100 | 7 | 1.4 | 3012 | PCT-US95-01087-1 | Sequence 1, Appl |
| | | | | US-08-811-566-2 | Sequence 2, Appl |

ALIGNMENTS

RESULT 1

US-09-198-452A-369
Sequence 369, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention of infection and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 369
LENGTH: 524
TYPE: PR
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-369

Query Match 60.8%; Score 313; DB 4; Length 524;
Best Local Similarity 99.6%; Pred. No. 3,98-239;
Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTKEEKPFGKLSFLMPHITHEKLVLMFMFCITFNTVLRDTRDILIVGAPSSGA 60
DB 10 MTKEEKPFGKLSFLMPHITHEKLVLMFMFCITFNTVLRDTRDILIVGAPSSGA 69
QY 61 EATPEIKFWLVVPCALIFMLIVAKLSNLSKQALFVAVGPPELIFPALPTVIVPLADVL 120
DB 70 EATPEIKFWLVVPCALIFMLIVAKLSNLSKQALFVAVGPPELIFPALPTVIVPLADVL 129
QY 121 HPTFEADRLQALIPPGILGLVAIIRNMTFAFYVLAELMGSVMLMFMGFANEITKHE 180
DB 130 HPTFEADRLQALIPPGILGLVAIIRNMTFAFYVLAELMGSVMLMFMGFANEITKHE 189
QY 181 AKRFYALFGIGANISILASGRAIYVASKLRASVSEGVDPWGISIRLIMAMTVSGVLVMA 240
DB 190 AKRFYALFGIGANISILASGRAIYVASKLRASVSEGVDPWGISIRLIMAMTVSGVLVMA 249
QY 241 SYMMINKNVLTDRFVYNEEMQKKGAKPRMNKDSFVLYDRSPYLLLTLLVIAVIGIC 300
DB 250 SYMMINKNVLTDRFVYNEEMQKKGAKPRMNKDSFVLYDRSPYLLLTLLVIAVIGIC 309
QY 301 INLEIETWKSQKLQYENNDYSEFMGNFSFMTGVSVLIMLFGVGNVIRKFGMLTGALV 360
DB 310 INLEIETWKSQKLQYENNDYSEFMGNFSFMTGVSVLIMLFGVGNVIRKFGMLTGALV 369
QY 361 TPVWVLLTGIIVFPALVIFRNQASGLVAMFGTTPMLAVVVGAIQNLISKSTKVALFDSTK 420
DB 370 TPVWVLLTGIIVFPALVIFRNQASGLVAMFGTTPMLAVVVGAIQNLISKSTKVALFDSTK 429
QY 421 EMAYIPIIDQOKYKGAIOVVARPKSGGALIQOGLVYCGSIGMTPTVLAIVILFII 480
DB 430 EMAYIPIIDQOKYKGAIOVVARPKSGGALIQOGLVYCGSIGMTPTVLAIVILFII 489
QY 481 AIMVLSATKLNKLFALQSALEKECEVAQEDSAPASS 515
DB 490 AIMVLSATKLNKLFALQSALEKECEVAQEDSAPASS 524

RESULT 2

US-09-328-352-5992
Sequence 5992, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 5992
LENGTH: 111
TYPE: PR
ORGANISM: Acinetobacter baumannii
US-09-328-352-5992

Query Match 1.6%; Score 8; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 5,5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 IVSGVLVM 239
DB 98 IVSGVLVM 105

RESULT 3

US-08-974-546-3
Sequence 3, Application US/08974546
Patent No. 5945287
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purni
TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Inocyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,546
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0428
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-845-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 137 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT18
CLONE: 3172266
US-08-974-546-3

Query Match 1.6%; Score 8; DB 2; Length 137;
Best Local Similarity 100.0%; Pred. No. 6,8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 SLLASGRA 202
DB 36 SLLASGRA 43

```
RESULT 4
US-09-252-991A-23710
; Sequence 23710, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23710
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23710

Query Match
1.6%; Score 8; DB 4; Length 172;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 LPPGLGL 140
Db 142 LPPGLGL 149

RESULT 5
5189147-3
; Patent No. 5189147
; APPLICANT: SAITO, HARUO; KRAZ, DAVID M.; EISEN, HERMAN N.;
; TONEGAWA, SUSUMU
; TITLE OF INVENTION: METEODIMERIC T LYMPHOCYTE RECEPTOR
; ANTIBODY
; NUMBER OF SEQUENCES: 21
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/271,216
; FILING DATE: 14-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 666,988
; FILING DATE: 31-OCT-1984
; APPLICATION NUMBER: 620,122
; FILING DATE: 13-JUN-1984
; SEQ ID NO:3
; LENGTH: 293
5189147-3

Query Match
1.6%; Score 8; DB 6; Length 293;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGLVMA 240
Db 279 VSGLVMA 286

RESULT 6
US-09-509-347-7
; Sequence 7, Application US/09509347
; Patent No. 6511830
; GENERAL INFORMATION:
; APPLICANT: KYOMA HAKKO KYOGYO CO., LTD.
; TITLE OF INVENTION: A KILLER T CELL RECEPTOR RECOGNIZING HUMAN
; FILE REFERENCE: PH-585PCT
; CURRENT APPLICATION NUMBER: US/09/509,347
; CURRENT FILING DATE: 2000-03-27
```

```
; PRIOR APPLICATION NUMBER: JP97/262536
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 7
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-509-347-7

Query Match
1.6%; Score 8; DB 4; Length 303;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGLVMA 240
Db 289 VSGLVMA 296

RESULT 7
US-09-384-212-2
; Sequence 2, Application US/09384212
; Patent No. 6177252
; GENERAL INFORMATION:
; APPLICANT: Lalouel, Jean-Marc
; APPLICANT: Rohwasser, Andreas
; TITLE OF INVENTION: Method to Determine Predisposition to Hypertension
; FILE REFERENCE: 2323-142
; CURRENT APPLICATION NUMBER: US/09/384,212
; CURRENT FILING DATE: 1999-08-27
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-384-212-2

Query Match
1.6%; Score 8; DB 3; Length 465;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 ADRLOAIL 133
Db 155 ADRLOAIL 162

RESULT 8
US-09-415-522-8
; Sequence 8, Application US/09415522A
; Patent No. 6291660
; GENERAL INFORMATION:
; APPLICANT: Gaffney, Thomas
; APPLICANT: Wendland, Juergen
; APPLICANT: Phillipsen, Peter
; TITLE OF INVENTION: No. 6291660el Fungal Genes Required For No. 6291660mal Growth And
; FILE REFERENCE: CGC2046
; CURRENT APPLICATION NUMBER: US/09/415,522A
; CURRENT FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Ashbya gossypii
US-09-415-522-8

Query Match
1.6%; Score 8; DB 3; Length 1013;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLASGR 201
```

Db 79 ISLASGR 86

RESULT 9

US-09-467-997-1
; Sequence 1, Application US/09467997
; Patent No. 6379925
; GENERAL INFORMATION:
; APPLICANT: Kitejewski, Jan
; APPLICANT: Uytendaele, Hendrik
; TITLE OF INVENTION: ANGIOGENIC MODULATION BY NOTCH SIGNAL TRANSDUCTION
; FILE REFERENCE: 53863-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/467,997
; CURRENT FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1964
; TYPE: PRT
; ORGANISM: mouse
US-09-467-997-1

Query Match 1.6%; Score 8; DB 4; Length 1964;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 GTTPLMLA 397
Db 1696 GTTPLMLA 1703

RESULT 10

US-07-894-063A-1
; Sequence 1, Application US/07894063A
; Patent No. 5980899
; GENERAL INFORMATION:
; APPLICANT: BERZOFSKY, Jay A.
; APPLICANT: SHIRAI, Mutsunori
; APPLICANT: AKATSUKA, Toshitaka
; APPLICANT: FEINSTONE, Stephen M.
; TITLE OF INVENTION: PEPTIDE FOR STIMULATION OF CYTOTOXIC T
; TITLE OF INVENTION: LYMPHOCYTES SPECIFIC FOR HEPATITIS C VIRUS IN A MAMMAL
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/894,063A
; FILING DATE: 19920610
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/162/NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear

US-07-894-063A-1

Query Match 1.4%; Score 7; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 TGAIVTP 362
Db 6 TGAIVTP 12

RESULT 11

US-07-894-063A-2
; Sequence 2, Application US/07894063A
; Patent No. 5980899
; GENERAL INFORMATION:
; APPLICANT: BERZOFSKY, Jay A.
; APPLICANT: SHIRAI, Mutsunori
; APPLICANT: AKATSUKA, Toshitaka
; APPLICANT: FEINSTONE, Stephen M.
; TITLE OF INVENTION: PEPTIDE FOR STIMULATION OF CYTOTOXIC T
; TITLE OF INVENTION: LYMPHOCYTES SPECIFIC FOR HEPATITIS C VIRUS IN A MAMMAL
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/894,063A
; FILING DATE: 19920610
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/162/NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear

US-07-894-063A-2

Query Match 1.4%; Score 7; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 TGAIVTP 362
Db 6 TGAIVTP 12

RESULT 12

US-09-393-634-11
; Sequence 11, Application US/09393634
; Patent No. 6558910
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark

APPLICANT: The Regents of the University of California
APPLICANT: The Government of the United States of America
APPLICANT: as represented by the Secretary of the
APPLICANT: Department of Health and Human Services
TITLE OF INVENTION: SF, a No. 6558910el Family of Taste Receptors
FILE REFERENCE: 02307E-0980005
CURRENT APPLICATION NUMBER: US/09/393,634
CURRENT FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 77
TYPE: PRT
ORGANISM: Mus sp.
FEATURE:
OTHER INFORMATION: mouse GR01
NAME/KEY: MOD_RES
LOCATION: (1)-(77)
OTHER INFORMATION: Xaa = any amino acid
US-09-393-634-11

Query Match 1.4%; Score 7; DB 4; Length 77;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 289 LITLTVI 295
DB 37 LITLTVI 43

RESULT 13
US-09-732-210-1020
Sequence 1020, Application US/09732210
Patent No. 6573361
GENERAL INFORMATION:
APPLICANT: Bunkers, Greg J.
APPLICANT: Liang, Jihong
APPLICANT: Mitlanck, Cindy A.
APPLICANT: Seale, Jeffrey W.
APPLICANT: Wu, Yonnie S.
TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
FILE REFERENCE: 38-21(15036)B
CURRENT APPLICATION NUMBER: US/09/732,210
CURRENT FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/169,513
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: US 60/169,340
PRIOR FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 1753
SEQ ID NO 1020
LENGTH: 110
TYPE: PRT
ORGANISM: Methanococcus jannaschii
US-09-732-210-1020

Query Match 1.4%; Score 7; DB 4; Length 110;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 82 YAKLSNI 88
DB 58 YAKLSNI 64

RESULT 14
US-08-444-818-44
Sequence 44, Application US/08444818
Patent No. 6150087
GENERAL INFORMATION:
APPLICANT: Chien, David Y.
APPLICANT: Rutter, William J.
TITLE OF INVENTION: NANBV Diagnostics and Vaccines
NUMBER OF SEQUENCES: 777

CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,818
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,590
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Hardin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0110.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-818-44

Query Match 1.4%; Score 7; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 356 TGAIVTP 362
DB 80 TGAIVTP 86

RESULT 15
US-08-444-818-52
Sequence 52, Application US/08444818
Patent No. 6150087
GENERAL INFORMATION:
APPLICANT: Chien, David Y.
APPLICANT: Rutter, William J.
TITLE OF INVENTION: NANBV Diagnostics and Vaccines
NUMBER OF SEQUENCES: 777
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,818
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,590
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Hardin, Alisa A.

```

;
;   REGISTRATION NUMBER: 33,895
;   REFERENCE/DOCKET NUMBER: 0110.002
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (508) 359-3876
;   TELEFAX: (508) 359-3885
;   INFORMATION FOR SEQ ID NO: 52:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH: 132 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;   MOLECULE TYPE: protein
;
US-08-444-818-52

Query Match      1.4%; Score 7; DB 3; Length 132;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      356 TGAIVTP 362
Db      57 TGAIVTP 63

RESULT 16
US-09-252-991A-29736
; Sequence 29736, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;   APPLICANT: Marc J. Rubenfield et al.
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;   FILE REFERENCE: 107196.136
;   CURRENT APPLICATION NUMBER: US/09/252,991A
;   CURRENT FILING DATE: 1999-02-18
;   PRIOR APPLICATION NUMBER: US 60/074,788
;   PRIOR FILING DATE: 1998-02-18
;   PRIOR APPLICATION NUMBER: US 60/094,190
;   PRIOR FILING DATE: 1998-07-27
;   NUMBER OF SEQ ID NOS: 33142
;   SEQ ID NO 29736
;   LENGTH: 159
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
;
US-09-252-991A-29736

Query Match      1.4%; Score 7; DB 4; Length 159;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      135 PGLGLV 141
Db      110 PGLGLV 116

RESULT 17
US-08-308-821A-6
; Sequence 6, Application US/08308821A
; Patent No. 5885567
; GENERAL INFORMATION:
;   APPLICANT: Sekellick, Margaret J.
;   APPLICANT: Marcus, Philip I.
;   APPLICANT: Ferrandino, Anthony F.
;   TITLE OF INVENTION: AVIAN INTERFERON GENES, NOVEL
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;   STREET: Two Militia Drive
;   CITY: Lexington
;   STATE: Massachusetts
;   COUNTRY: USA
;   ZIP: 02173
;
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
```

```

;
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/308,821A
;   FILING DATE: 19-SEP-1994
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/235,402
;   FILING DATE: 28-APR-1994
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/139,909
;   FILING DATE: 22-OCT-1993
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Carroll, Alice O.
;   REGISTRATION NUMBER: 33,542
;   REFERENCE/DOCKET NUMBER: UCT93-04A2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 861-6240
;   TELEFAX: (617) 861-9540
;   INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH: 193 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;   MOLECULE TYPE: protein
;
US-08-308-821A-6

Query Match      1.4%; Score 7; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      287 ILTLTL 293
Db      15 ILTLTL 21

RESULT 18
US-08-831-627-6
; Sequence 6, Application US/08831627
; Patent No. 6020465
; GENERAL INFORMATION:
;   APPLICANT: Sekellick, Margaret J.
;   APPLICANT: Marcus, Philip I.
;   APPLICANT: Ferrandino, Anthony F.
;   TITLE OF INVENTION: CHICKEN INTERFERON GENE AND NOVEL
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;   STREET: Two Militia Drive
;   CITY: Lexington
;   STATE: Massachusetts
;   COUNTRY: USA
;   ZIP: 02173
;
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/831,627
FILING DATE: 09-APR-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/235,402
FILING DATE: 28-APR-1994
APPLICATION NUMBER: US 08/139,909
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: UCT93-04A
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 193 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-831-627-6

Query Match 1.4%; Score 7; DB 3; Length 193;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILLLTLL 293
DB 15 ILLLTLL 21

RESULT 19
US-08-765-381-15
Sequence 15; Application US/08765381
Patent No. 6083724
GENERAL INFORMATION:
APPLICANT: Commonwealth Scientific and Industrial Research Organisation
TITLE OF INVENTION: No. 6083724el avian cytokines and genetic
TITLE OF INVENTION: sequences encoding same
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Scully Scott Murphy and Presser
STREET: 400 Garden City Plaza
CITY: Garden City, New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/765,381
FILING DATE: 19-DEC-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: AU PN1542/95
FILING DATE: 06-MAR-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/AU96/00114
FILING DATE: 05-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: PRESSER, LEOPOLD
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-516-742-4343
TELEFAX: 1-516-742-4366
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 193 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: avian
IMMEDIATE SOURCE:
CLONE: ChIFN-alpha
US-08-765-381-15

Query Match 1.4%; Score 7; DB 3; Length 193;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILLLTLL 293
DB 15 ILLLTLL 21

RESULT 20
US-08-804-372A-32
Sequence 32; Application US/08804372A
Patent No. 6183753
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
APPLICANT: Wild, Martha A.
APPLICANT: Winslow, Barbara J.
TITLE OF INVENTION: Recombinant Chimeric Viruses and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,372A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 2552/39115E
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 197 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-804-372A-32

Query Match 1.4%; Score 7; DB 3; Length 197;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILLLTLL 293
DB 15 ILLLTLL 21

RESULT 21
PCT-US94-04174-18
Sequence 18; Application PC/TUS9404174
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory
APPLICANT: Kim, Jungshuh P.
APPLICANT: Moeckli, Randolph
TITLE OF INVENTION: HCV Immunodiagnostic Antigens and Antibodies
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Law Offices of Peter J. Dehlinger
STREET: 350 Cambridge Avenue, Suite 300
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94306

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04174
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 505,611
FILING DATE: 06-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0107
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-324-0880
TELEFAX: 415-324-0960
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 211 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Clone 36-2 protein sequence
PCT-US94-04174-18

Query Match 1.4%; Score 7; DB 5; Length 211;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 TGAIVTP 362
|||
Db 142 TGAIVTP 148

RESULT 22
US-09-252-991A-21062
Sequence 21062, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR APPLICATION DATA:
FILING DATE: 1999-02-18
APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 21062
LENGTH: 216
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21062

Query Match 1.4%; Score 7; DB 4; Length 216;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 381 QASGLVA 387
|||
Db 113 QASGLVA 119

RESULT 23
US-09-328-352-4385

Sequence 4385, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4385
LENGTH: 218
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-4385

Query Match 1.4%; Score 7; DB 4; Length 218;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 474 VILLFTI 480
|||
Db 50 VILLFTI 56

RESULT 24
US-09-107-532A-7302
Sequence 7302, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
City: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 893-5007
TELEFAX: (781) 893-8277
INFORMATION FOR SEQ ID NO: 7302:
SEQUENCE CHARACTERISTICS:
LENGTH: 225 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...225

SEQUENCE DESCRIPTION: SEQ ID NO: 7302:
US-09-107-532A-7302

Query Match 1.4%; Score 7; DB 4; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 RDTKDTL 51
| | | | |
| | | | |
Db 76 RDTKDTL 82

RESULT 25
US-09-333-599-2
; Sequence 2, Application US/09333599
; Patent No. 6245898
; GENERAL INFORMATION:
; APPLICANT: Testa, Jacqueline E.
; APPLICANT: Quigley, James P.
; APPLICANT: Seandel, Marco
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS
; FILE REFERENCE: SUNY
; CURRENT APPLICATION NUMBER: US/09/333,599
; CURRENT FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-333-599-2

Query Match 1.4%; Score 7; DB 3; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLLASG 200
| | | | |
| | | | |
Db 49 ISLLASG 55

RESULT 26
US-09-499-781-2
; Sequence 2, Application US/09499781
; Patent No. 6498014
; GENERAL INFORMATION:
; APPLICANT: Testa, Jacqueline E.
; APPLICANT: Quigley, James P.
; APPLICANT: Seandel, Marco
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS
; FILE REFERENCE: SUNY
; CURRENT APPLICATION NUMBER: US/09/499,781
; CURRENT FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 09/333,599
; PRIOR FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-499-781-2

Query Match 1.4%; Score 7; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLLASG 200
| | | | |
| | | | |
Db 49 ISLLASG 55

RESULT 27
US-09-252-991A-19115
; Sequence 19115, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19115
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19115

Query Match 1.4%; Score 7; DB 4; Length 259;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 IYFFALV 376
| | | | |
| | | | |
Db 55 IYFFALV 61

RESULT 28
US-09-328-352-7144
; Sequence 7144, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7144
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7144

Query Match 1.4%; Score 7; DB 4; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 NISLAS 199
| | | | |
| | | | |
Db 239 NISLAS 245

RESULT 29
US-09-414-276-3
; Sequence 3, Application US/09414276
; Patent No. 6392121
; GENERAL INFORMATION:
; APPLICANT: Mason, Hugh
; APPLICANT: Palmer, Kenneth
; APPLICANT: Heferson, Kathleen
; APPLICANT: Mor, Tsafir
; APPLICANT: Arntzen, Charles
; TITLE OF INVENTION: Gemini Virus Vectors for Gene Expression in Plants
; FILE REFERENCE: 4868/84453
; CURRENT APPLICATION NUMBER: US/09/414,276
; CURRENT FILING DATE: 1999-10-07

NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 293
TYPE: PRT
ORGANISM: bean yellow dwarf virus
US-09-414-276-3

Query Match 1.4%; Score 7; DB 4; Length 293;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 100 TPFLIF 106
Db 39 TPFLIF 45

RESULT 30
US-09-134-001C-4537
Sequence 4537, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4537
LENGTH: 317
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4537

Query Match 1.4%; Score 7; DB 4; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 ELKKVLP 29
Db 46 ELKKVLP 52

RESULT 31
US-09-414-276-6
Sequence 6, Application US/09414276
Patent No. 6392121
GENERAL INFORMATION:
APPLICANT: Mason, Hugh
APPLICANT: Palmer, Kenneth
APPLICANT: Hefferon, Kathleen
APPLICANT: Mor, Tsafir
APPLICANT: Arntzen, Charles
TITLE OF INVENTION: Gemini Virus Vectors for Gene Expression in Plants
FILE REFERENCE: 4868/84453
CURRENT APPLICATION NUMBER: US/09/414,276
CURRENT FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 335
TYPE: PRT
ORGANISM: bean yellow dwarf virus
US-09-414-276-6

Query Match 1.4%; Score 7; DB 4; Length 335;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 100 TPFLIF 106
Db 39 TPFLIF 45

RESULT 32
US-09-107-532A-5132
Sequence 5132, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5132:
SEQUENCE CHARACTERISTICS:
LENGTH: 357 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...357
SEQUENCE DESCRIPTION: SEQ ID NO: 5132:
US-09-107-532A-5132

Query Match 1.4%; Score 7; DB 4; Length 357;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 IVEFALV 376
Db 21 IVEFALV 27

RESULT 33
US-09-252-991A-18068
Sequence 18068, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

;; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;; FILE REFERENCE: 107196.136
;; CURRENT APPLICATION NUMBER: US/09/252.991A
;; CURRENT FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 18068
;; LENGTH: 362
;; TYPE: PRT
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18068

Query Match 1.4%; Score 7; DB 4; Length 362;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 PFLIFPA 107
DB 177 PFLIFPA 183

RESULT 34
US-09-622-439-2
; Sequence 2; Application US/09622439
; Patent No. 6553344
; GENERAL INFORMATION:
; APPLICANT: Yamanouchi Pharmaceutical Co., Ltd.
; TITLE OF INVENTION: A novel G protein coupled receptor protein
; FILE REFERENCE: Y9905
; CURRENT APPLICATION NUMBER: US/09/622,439
; CURRENT FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: JP P1998-060245
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: JP P1999-026774
; PRIOR FILING DATE: 1999-02-03
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 375
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-622-439-2

Query Match 1.4%; Score 7; DB 4; Length 375;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 LAVVGA 402
DB 190 LAVVGA 196

RESULT 35
US-09-107-532A-6371
; Sequence 6371, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Walham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660

;; COMPUTER: PC
;; OPERATING SYSTEM: <Unknown>
;; SOFTWARE: ASCII
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/107,532A
;; FILING DATE: 30-Jun-1998
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 60/085,598
;; FILING DATE: 14 May 1998
;; APPLICATION NUMBER: 60/051571
;; FILING DATE: July 2, 1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Arinello, Pamela Deneka
;; REGISTRATION NUMBER: 40,489
;; REFERENCE/DOCKET NUMBER: GTC-012
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (781)893-5007
;; TELEFAX: (781)893-8277
;; INFORMATION FOR SEQ ID NO: 6371:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 414 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: YES
;; ORIGINAL SOURCE:
;; ORGANISM: Enterococcus faecium
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (B) LOCATION 1...414
;; SEQUENCE DESCRIPTION: SEQ ID NO: 6371:
US-09-107-532A-6371

Query Match 1.4%; Score 7; DB 4; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GLVIMAS 241
DB 4 GLVIMAS 10

RESULT 36
US-09-328-352-6097
; Sequence 6097, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6097
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6097

Query Match 1.4%; Score 7; DB 4; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 ANISLA 198
DB 159 ANISLA 165

RESULT 37
US-09-328-352-7964
; Sequence 7964, Application US/09328352
; Patent No. 6562958

GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS
CURRENT APPLICATION NUMBER: GTC99-03PA
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 7964
LENGTH: 425
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-7964

Query Match 1.4%; Score 7; DB 4; Length 425;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 AEAIPFI 66
Db 65 AEAIPFI 71

RESULT 38
US-09-252-991A-20578
Sequence 20578, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
CURRENT APPLICATION NUMBER: 107196.136
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 20578
LENGTH: 452
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20578

Query Match 1.4%; Score 7; DB 4; Length 452;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 SGAEAP 64
Db 436 SGAEAP 442

RESULT 39
US-07-853-985A-6
Sequence 6, Application US/07853985A
Patent No. 5436318
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory
APPLICANT: Kim, Jungshuh P.
APPLICANT: Moeckli, Randolph
TITLE OF INVENTION: Hepatitis C Virus Epitopes
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Peter J. Dehlinger
STREET: P.O. BOX 60850
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/853,985A
FILING DATE: 19920320
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/505,611
FILING DATE: 06-APR-1990
APPLICATION NUMBER: US 07/594,854
FILING DATE: 09-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Dehlinger, Peter J.
REGISTRATION NUMBER: 28,006
REFERENCE/DOCKET NUMBER: 4600-0076.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-324-0880
TELEFAX: 415-324-0960
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 504 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-853-985A-6

Query Match 1.4%; Score 7; DB 1; Length 504;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 356 TGAIVTP 362
Db 308 TGAIVTP 314

RESULT 40
US-07-681-703B-6
Sequence 6, Application US/07681703B
Patent No. 543965
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory
APPLICANT: Kim, Jungshuh P.
APPLICANT: Moeckli, Randolph
TITLE OF INVENTION: Hepatitis C Virus Epitopes
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Ave., Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/681,703B
FILING DATE: 05-APR-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 505,611
FILING DATE: 06-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 594,854
FILING DATE: 09-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-076.21
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-0880
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 504 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-681-703B-6

Query Match 1.4%; Score 7; DB 1; Length 504;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 356 TGAIVTP 362 |||||
DB 308 TGAIVTP 314

Search completed: November 25, 2003, 10:08:02
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:07:06 ; Search time 55 Seconds
(without alignments)
1727.060 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 515

Sequence: 1 MKRTEKPKPKGLRSFLMPH.....AQSALKEQVADSDAPASS 515

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 673684 seqs, 18443283 residues

Word size : 0

Total number of hits satisfying chosen parameters: 673684

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|------------------------|----------------------|
| 1 | 515 | 100.0 | 515 | 9 US-09-892-851-2 | Sequence 2, Appl 1 |
| 2 | 8 | 1.6 | 73 | 12 US-10-029-386-28700 | Sequence 28700, A |
| 3 | 8 | 1.6 | 302 | 9 US-09-789-6978-21 | Sequence 21, Appl 1 |
| 4 | 8 | 1.6 | 303 | 12 US-09-795-271-57 | Sequence 57, Appl 1 |
| 5 | 8 | 1.6 | 310 | 12 US-09-965-422-47 | Sequence 47, Appl 1 |
| 6 | 8 | 1.6 | 312 | 12 US-10-017-161-508 | Sequence 508, Appl 1 |
| 7 | 8 | 1.6 | 312 | 12 US-09-965-422-46 | Sequence 46, Appl 1 |
| 8 | 8 | 1.6 | 312 | 12 US-09-965-422-48 | Sequence 48, Appl 1 |
| 9 | 8 | 1.6 | 316 | 12 US-09-795-271-56 | Sequence 56, Appl 1 |
| 10 | 8 | 1.6 | 316 | 12 US-09-912-976-64 | Sequence 64, Appl 1 |
| 11 | 8 | 1.6 | 316 | 12 US-09-965-422-45 | Sequence 45, Appl 1 |
| 12 | 8 | 1.6 | 666 | 15 US-10-225-630-10 | Sequence 10, Appl 1 |
| 13 | 7 | 1.4 | 20 | 14 US-10-032-482-23 | Sequence 23, Appl 1 |
| 14 | 7 | 1.4 | 48 | 9 US-09-864-761-36243 | Sequence 36243, A |
| 15 | 7 | 1.4 | 59 | 9 US-09-764-853-614 | Sequence 614, Appl 1 |

| | | | | | |
|----|---|-----|------|------------------------|-----------------------|
| 16 | 7 | 1.4 | 59 | 9 US-09-764-898-271 | Sequence 271, Appl 1 |
| 17 | 7 | 1.4 | 59 | 15 US-10-073-865-100 | Sequence 100, Appl 1 |
| 18 | 7 | 1.4 | 77 | 9 US-09-393-634-11 | Sequence 11, Appl 1 |
| 19 | 7 | 1.4 | 77 | 12 US-10-383-982-11 | Sequence 11, Appl 1 |
| 20 | 7 | 1.4 | 98 | 15 US-10-144-929-125 | Sequence 125, Appl 1 |
| 21 | 7 | 1.4 | 106 | 15 US-10-032-482-3 | Sequence 3, Appl 1 |
| 22 | 7 | 1.4 | 135 | 11 US-09-991-936-38 | Sequence 38, Appl 1 |
| 23 | 7 | 1.4 | 135 | 11 US-09-991-936-41 | Sequence 41, Appl 1 |
| 24 | 7 | 1.4 | 137 | 10 US-09-921-397-106 | Sequence 106, Appl 1 |
| 25 | 7 | 1.4 | 144 | 12 US-10-002-631C-6 | Sequence 6, Appl 1 |
| 26 | 7 | 1.4 | 174 | 12 US-10-017-161-700 | Sequence 700, Appl 1 |
| 27 | 7 | 1.4 | 181 | 9 US-09-925-297-778 | Sequence 778, Appl 1 |
| 28 | 7 | 1.4 | 202 | 10 US-09-738-626-6137 | Sequence 6137, Appl 1 |
| 29 | 7 | 1.4 | 209 | 15 US-10-153-668-424 | Sequence 424, Appl 1 |
| 30 | 7 | 1.4 | 209 | 15 US-10-153-668-478 | Sequence 478, Appl 1 |
| 31 | 7 | 1.4 | 234 | 12 US-10-238-075-643 | Sequence 643, Appl 1 |
| 32 | 7 | 1.4 | 253 | 15 US-10-153-668-314 | Sequence 314, Appl 1 |
| 33 | 7 | 1.4 | 253 | 15 US-10-153-668-476 | Sequence 476, Appl 1 |
| 34 | 7 | 1.4 | 253 | 16 US-10-156-136-35 | Sequence 35, Appl 1 |
| 35 | 7 | 1.4 | 282 | 9 US-09-925-301-1234 | Sequence 1234, Appl 1 |
| 36 | 7 | 1.4 | 283 | 15 US-10-151-336-3 | Sequence 3, Appl 1 |
| 37 | 7 | 1.4 | 306 | 9 US-09-835-788A-19 | Sequence 19, Appl 1 |
| 38 | 7 | 1.4 | 306 | 12 US-10-175-042-19 | Sequence 19, Appl 1 |
| 39 | 7 | 1.4 | 321 | 15 US-10-146-772-12 | Sequence 12, Appl 1 |
| 40 | 7 | 1.4 | 329 | 15 US-10-208-018-104 | Sequence 104, Appl 1 |
| 41 | 7 | 1.4 | 335 | 15 US-10-151-336-6 | Sequence 6, Appl 1 |
| 42 | 7 | 1.4 | 336 | 12 US-10-238-075-305 | Sequence 305, Appl 1 |
| 43 | 7 | 1.4 | 339 | 9 US-09-729-674-138 | Sequence 138, Appl 1 |
| 44 | 7 | 1.4 | 352 | 11 US-09-925-299-963 | Sequence 963, Appl 1 |
| 45 | 7 | 1.4 | 352 | 11 US-09-925-299-963 | Sequence 963, Appl 1 |
| 46 | 7 | 1.4 | 364 | 9 US-09-815-242-13652 | Sequence 13652, A |
| 47 | 7 | 1.4 | 374 | 15 US-10-255-841-10 | Sequence 10, Appl 1 |
| 48 | 7 | 1.4 | 375 | 12 US-10-272-983-16 | Sequence 16, Appl 1 |
| 49 | 7 | 1.4 | 375 | 12 US-10-393-807-16 | Sequence 16, Appl 1 |
| 50 | 7 | 1.4 | 375 | 15 US-10-318-142-2 | Sequence 2, Appl 1 |
| 51 | 7 | 1.4 | 375 | 15 US-10-225-567A-494 | Sequence 494, Appl 1 |
| 52 | 7 | 1.4 | 393 | 12 US-10-091-007A-234 | Sequence 234, Appl 1 |
| 53 | 7 | 1.4 | 404 | 16 US-10-080-170-332 | Sequence 332, Appl 1 |
| 54 | 7 | 1.4 | 437 | 12 US-10-032-588-7130 | Sequence 7130, Appl 1 |
| 55 | 7 | 1.4 | 459 | 15 US-10-156-761-14298 | Sequence 14298, A |
| 56 | 7 | 1.4 | 466 | 12 US-10-238-075-480 | Sequence 480, Appl 1 |
| 57 | 7 | 1.4 | 481 | 12 US-09-882-227-134 | Sequence 134, Appl 1 |
| 58 | 7 | 1.4 | 487 | 15 US-10-156-761-14234 | Sequence 14234, A |
| 59 | 7 | 1.4 | 487 | 15 US-10-106-698-4685 | Sequence 4685, Appl 1 |
| 60 | 7 | 1.4 | 523 | 15 US-10-156-761-11594 | Sequence 11594, A |
| 61 | 7 | 1.4 | 541 | 12 US-10-181-157-1 | Sequence 1, Appl 1 |
| 62 | 7 | 1.4 | 542 | 12 US-10-181-157-1 | Sequence 1, Appl 1 |
| 63 | 7 | 1.4 | 579 | 15 US-10-011-548-2 | Sequence 2, Appl 1 |
| 64 | 7 | 1.4 | 590 | 15 US-09-733-183A-4 | Sequence 4, Appl 1 |
| 65 | 7 | 1.4 | 590 | 12 US-10-342-372-4 | Sequence 4, Appl 1 |
| 66 | 7 | 1.4 | 591 | 12 US-09-733-183A-2 | Sequence 2, Appl 1 |
| 67 | 7 | 1.4 | 591 | 12 US-10-342-372-2 | Sequence 2, Appl 1 |
| 68 | 7 | 1.4 | 593 | 12 US-10-241-827-2 | Sequence 2, Appl 1 |
| 69 | 7 | 1.4 | 604 | 10 US-09-862-027-17 | Sequence 17, Appl 1 |
| 70 | 7 | 1.4 | 606 | 15 US-10-156-761-9993 | Sequence 9993, Appl 1 |
| 71 | 7 | 1.4 | 625 | 10 US-09-738-626-6787 | Sequence 6787, Appl 1 |
| 72 | 7 | 1.4 | 686 | 12 US-10-080-608A-33 | Sequence 33, Appl 1 |
| 73 | 7 | 1.4 | 686 | 12 US-10-370-688-122 | Sequence 122, Appl 1 |
| 74 | 7 | 1.4 | 760 | 12 US-10-238-075-943 | Sequence 943, Appl 1 |
| 75 | 7 | 1.4 | 786 | 11 US-09-950-041-2 | Sequence 2, Appl 1 |
| 76 | 7 | 1.4 | 884 | 12 US-10-179-766-10 | Sequence 10, Appl 1 |
| 77 | 7 | 1.4 | 1000 | 10 US-09-929-955-9 | Sequence 9, Appl 1 |
| 78 | 7 | 1.4 | 1050 | 15 US-10-104-966-9 | Sequence 9, Appl 1 |
| 79 | 7 | 1.4 | 1050 | 15 US-10-156-761-8816 | Sequence 8816, Appl 1 |
| 80 | 7 | 1.4 | 1199 | 15 US-10-156-761-10084 | Sequence 10084, A |
| 81 | 7 | 1.4 | 1884 | 10 US-09-785-770A-17 | Sequence 17, Appl 1 |
| 82 | 7 | 1.4 | 1907 | 10 US-09-785-770A-16 | Sequence 16, Appl 1 |
| 83 | 7 | 1.4 | 3011 | 9 US-09-742-659-4 | Sequence 4, Appl 1 |
| 84 | 7 | 1.4 | 3011 | 9 US-09-916-359-2 | Sequence 2, Appl 1 |
| 85 | 7 | 1.4 | 3011 | 10 US-09-238-076-20 | Sequence 20, Appl 1 |
| 86 | 7 | 1.4 | 3011 | 10 US-09-952-572-9 | Sequence 9, Appl 1 |
| 87 | 7 | 1.4 | 3011 | 10 US-09-929-955-1 | Sequence 1, Appl 1 |
| 88 | 7 | 1.4 | 3011 | 10 US-09-747-419-20 | Sequence 20, Appl 1 |

```

89 7 1.4 3011 11 US-09-891-894-3 Sequence 3, Appl1
90 7 1.4 3011 11 US-09-995-937-20 Sequence 20, Appl1
91 7 1.4 3011 11 US-09-917-563-20 Sequence 20, Appl1
92 7 1.4 3011 12 US-10-184-150-3 Sequence 3, Appl1
93 7 1.4 3011 14 US-10-104-966-1 Sequence 1, Appl1
94 7 1.4 3011 15 US-10-259-275-20 Sequence 2, Appl1
95 7 1.4 3011 16 US-10-232-643-6 Sequence 6, Appl1
96 7 1.4 3012 10 US-09-238-076-2 Sequence 2, Appl1
97 7 1.4 3012 11 US-09-995-937-2 Sequence 2, Appl1
98 7 1.4 3012 11 US-09-917-563-2 Sequence 109, App
99 7 1.4 3562 12 US-10-341-434-109 Sequence 11, Appl1
100 6 1.2 8 10 US-09-910-071-11

```

ALIGNMENTS

```

RESULT 1
US-09-892-851-2
; Sequence 2, Application US/09892851
; Patent No. US20020081682A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pasteur Limited
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
; FILE REFERENCE: 77813-6
; CURRENT APPLICATION NUMBER: US/09/892,851
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/114,060
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: US 60/123,967
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 60/144,271
; PRIOR FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-892-851-2

```

```

Query Match 100.0%; Score 515; DB 9; Length 515;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKTEERPGKLSFLMPHITHLKKVLPNFMFCITFNTVTRDKDTLIVGAPSSGA 60
Db 1 MTKTEERPGKLSFLMPHITHLKKVLPNFMFCITFNTVTRDKDTLIVGAPSSGA 60
QY 61 EALPFIKFWLVVPCAIIFMLIYAKLSNLSKQALFYAVGPELIFPFLPTVYPLRDVL 120
Db 61 EALPFIKFWLVVPCAIIFMLIYAKLSNLSKQALFYAVGPELIFPFLPTVYPLRDVL 120
QY 121 HPTFEADRLQALIPPGILGLVALIRNMTFAFYVLAELMGSVMLSMFWGFANEITKHE 180
Db 121 HPTFEADRLQALIPPGILGLVALIRNMTFAFYVLAELMGSVMLSMFWGFANEITKHE 180
QY 181 AKRFYALFGIGANTSLASGRAIYWASKLRASVEGVDPWGISRLIMAMTVYSGVLVMA 240
Db 181 AKRFYALFGIGANTSLASGRAIYWASKLRASVEGVDPWGISRLIMAMTVYSGVLVMA 240
QY 241 SYMMINKNVLTDPFYAPBEKMGKKGAKPMNNKDSFLYDRSPYLLLTLLVIAVIGIC 300
Db 241 SYMMINKNVLTDPFYAPBEKMGKKGAKPMNNKDSFLYDRSPYLLLTLLVIAVIGIC 300
QY 301 INLEIETWKSQKLQYFNNDYSEFGNFSFTGVSVLIMLFVGNVIRKFGWLIGALV 360
Db 301 INLEIETWKSQKLQYFNNDYSEFGNFSFTGVSVLIMLFVGNVIRKFGWLIGALV 360
QY 361 TPVWVLLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVVGAIQNIISKSTKIALPSTK 420
Db 361 TPVWVLLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVVGAIQNIISKSTKIALPSTK 420

```

```

QY 421 EMAYIPLDQKQKGAIDVVAARFGSGGALIQGLVITCGSIGAMTPYLAIVLFI 480
Db 421 EMAYIPLDQKQKGAIDVVAARFGSGGALIQGLVITCGSIGAMTPYLAIVLFI 480
QY 481 AIWLVSATKLNKFLAQSALKEQVAGDSAPASS 515
Db 481 AIWLVSATKLNKFLAQSALKEQVAGDSAPASS 515

```

```

RESULT 2
US-10-029-386-28700
; Sequence 28700, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEWICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28700
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR6.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 16
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 15
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 11
; OTHER INFORMATION: SWISSPROT HIT: Q9GZK4, EVALU 1.00e-36
US-10-029-386-28700

```

```

Query Match 1.6%; Score 8; DB 12; Length 73;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
Db 42 LFYAVGTP 49

RESULT 3
US-09-789-697A-21
; Sequence 21, Application US/09789697A
; Patent No. US20020064521A1
; GENERAL INFORMATION:
; APPLICANT: Ellenborn, Joshua D.I.
; APPLICANT: Diamond, Don J.
; TITLE OF INVENTION: p53-Specific T Cell Receptor for Adoptive Immunotherapy
; FILE REFERENCE: 1954-279-11
; CURRENT APPLICATION NUMBER: US/09/789,697A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 60/183,752
; PRIOR FILING DATE: 2000-02-22
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-789-697A-21

Query Match 1.6%; Score 8; DB 9; Length 302;
Best Local Similarity 100.0%; Pred. No. 28;

```

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGLVMA 240
|||||

Db 288 VSGLVMA 295

RESULT 4
US-09-795-271-57

Sequence 57, Application US/09795271
Publication No. US20030165829A1
GENERAL INFORMATION:
APPLICANT: Padigaru, Muralidhara
APPLICANT: Majumder, Kumud
APPLICANT: Burgess, Catherine E
APPLICANT: Vermet, Corine A.M
APPLICANT: Fernandes, Elma
APPLICANT: Shimkets, Richard A
APPLICANT: Tchernev, Velizar T
APPLICANT: Mishra, Vishnu
APPLICANT: Casman, Stacie
APPLICANT: Spytek, Kimberly A
APPLICANT: Zerhusen, Bryan
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 15966-694
CURRENT APPLICATION NUMBER: US/09/795,271
CURRENT FILING DATE: 2001-02-27
PRIOR APPLICATION NUMBER: 60/185,674
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: 60/185,535
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: 60/186,585
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,604
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,584
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,717
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,716
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,719
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,827
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/218,323
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/218,435
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/220,517
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: 60/223,897
PRIOR FILING DATE: 2000-08-09
PRIOR APPLICATION NUMBER: 60/260,020
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/264,849
PRIOR FILING DATE: 2001-01-29
PRIOR APPLICATION NUMBER: 60/186,715
PRIOR FILING DATE: 2000-03-03
NUMBER OF SEQ ID NOS: 83
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 57
LENGTH: 303
TYPE: PRT
ORGANISM: Homo sapiens
US-09-795-271-57

Query Match 1.6%; Score 8; DB 12; Length 303;
Best Local Similarity 100.0%; Pred. No. 28;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
|||||

Db 265 LFYAVGTP 272

RESULT 5
US-09-965-422-47

Sequence 47, Application US/09965422
Publication No. US20030216545A1
GENERAL INFORMATION:
APPLICANT: Spytek, Kimberly A
APPLICANT: Casman, Stacie
APPLICANT: Padigaru, Muralidhara
APPLICANT: Dickson, Kevin
APPLICANT: Vermet, Corine
APPLICANT: Spaderma, Steven K
APPLICANT: Shenoy, Suresh G
APPLICANT: Gerlach, Valerie
APPLICANT: Ellerman, Karen
APPLICANT: Edinger, Shlomit
APPLICANT: MacDougall, John R
APPLICANT: Smithson, Glenda
APPLICANT: Li, Li
APPLICANT: Malvankar, Uriel M
APPLICANT: Taylor, Sarah
APPLICANT: Gunther, Erik
APPLICANT: Tchernev, Velizar T
TITLE OF INVENTION: No. US20030216545A1el Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21401-132
CURRENT APPLICATION NUMBER: US/09/965,422
CURRENT FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/236,286
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: 60/236,284
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: 60/237,581
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/238,735
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/240,736
PRIOR FILING DATE: 2000-10-16
PRIOR APPLICATION NUMBER: 60/260,019
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/260,338
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 60/262,156
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 60/262,498
PRIOR FILING DATE: 2001-01-18
PRIOR APPLICATION NUMBER: 60/263,133
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/263,691
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/266,109
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/271,634
PRIOR FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 127
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 47
LENGTH: 310
TYPE: PRT
ORGANISM: Mus musculus
US-09-965-422-47

Query Match 1.6%; Score 8; DB 12; Length 310;
Best Local Similarity 100.0%; Pred. No. 29;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
|||||

Db 274 LFYAVGTP 281

RESULT 6

```
US-10-017-161-508
; Sequence 508, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUMA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKUTAMA, YUTAKA
; APPLICANT: ARIYATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 508
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-508

Query Match      1.6%; Score 8; DB 12; Length 312;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      94 LFYAVGTP 101
Db      274 LFYAVGTP 281

RESULT 7
US-09-965-422-46
; Sequence 46, Application US/09965422
; Publication No. US20030216545A1
; GENERAL INFORMATION:
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Casman, Stacie
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Dickson, Kevin
; APPLICANT: Vernet, Corine
; APPLICANT: Spaderna, Steven K
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Gerlach, Valerie
; APPLICANT: Ellerman, Karen
; APPLICANT: Edinger, Shlomit
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenda
; APPLICANT: Li, Li
; APPLICANT: Malyankar, Urial M
; APPLICANT: Taylor, Sarah
; APPLICANT: Gunther, Erik
; APPLICANT: Tcherenev, Velizar T
; TITLE OF INVENTION: No. US20030216545A1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21401-132
; CURRENT APPLICATION NUMBER: US/09/965,422
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/236,286
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/236,284
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/237,581
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/238,735
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/240,736
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/260,019
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: 60/260,338
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/262,156
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/263,133
; PRIOR FILING DATE: 2001-01-19

US-09-965-422-48
; Sequence 48, Application US/09965422
; Publication No. US20030216545A1
; GENERAL INFORMATION:
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Casman, Stacie
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Dickson, Kevin
; APPLICANT: Vernet, Corine
; APPLICANT: Spaderna, Steven K
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Gerlach, Valerie
; APPLICANT: Ellerman, Karen
; APPLICANT: Edinger, Shlomit
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenda
; APPLICANT: Li, Li
; APPLICANT: Malyankar, Urial M
; APPLICANT: Taylor, Sarah
; APPLICANT: Gunther, Erik
; APPLICANT: Tcherenev, Velizar T
; TITLE OF INVENTION: No. US20030216545A1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21401-132
; CURRENT APPLICATION NUMBER: US/09/965,422
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/236,286
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/236,284
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/237,581
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/238,735
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/240,736
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/260,019
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: 60/260,338
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/262,156
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/262,498
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/263,133
; PRIOR FILING DATE: 2001-01-19
```

PRIOR APPLICATION NUMBER: 60/263,691
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/266,109
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/271,634
PRIOR FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 127
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 48
LENGTH: 312
TYPE: PRT
ORGANISM: Mus musculus
US-09-965-422-48

Query Match 1.6%; Score 8; DB 12; Length 312;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
Db 274 LFYAVGTP 281

RESULT 9
US-09-795-271-56
Sequence 56, Application US/09795271
Publication No. US20030165829A1
GENERAL INFORMATION:
APPLICANT: Padigar, Muralidhara
APPLICANT: Majumder, Kumud
APPLICANT: Burgess, Catherine E
APPLICANT: Vernet, Corine A.M
APPLICANT: Fernandes, Elma
APPLICANT: Shimkets, Richard A
APPLICANT: Tcherenev, Velizar T
APPLICANT: Mishra, Vishnu
APPLICANT: Caeman, Stacie
APPLICANT: Spyttek, Kimberly A
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 15966-694
CURRENT FILING DATE: 2001-02-27
PRIOR APPLICATION NUMBER: 60/185,674
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: 60/185,535
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: 60/186,585
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,604
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,584
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,717
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,716
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,719
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,827
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/218,323
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/218,435
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/220,517
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: 60/223,897
PRIOR FILING DATE: 2000-08-09
PRIOR APPLICATION NUMBER: 60/260,020
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/264,849
PRIOR FILING DATE: 2001-01-29

PRIOR APPLICATION NUMBER: 60/186,715
PRIOR FILING DATE: 2000-03-03
NUMBER OF SEQ ID NOS: 83
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 56
LENGTH: 316
TYPE: PRT
ORGANISM: Homo sapiens
US-09-795-271-56

Query Match 1.6%; Score 8; DB 12; Length 316;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
Db 274 LFYAVGTP 281

RESULT 10
US-09-912-976-64
Sequence 64, Application US/09912976
Publication No. US20030212255A1
GENERAL INFORMATION:
APPLICANT: Padigar, Muralidhara
APPLICANT: Mezes, Peter
APPLICANT: Burgess, Catherine
APPLICANT: Caeman, Stacie
APPLICANT: Grose, William M
APPLICANT: Alebrook II, John P
APPLICANT: Lepley, Denise M
APPLICANT: Gerlach, Valerie L
APPLICANT: MacDougall, John R
APPLICANT: Smithson, Glenda
APPLICANT: Mishra, Vishnu
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-070
CURRENT APPLICATION NUMBER: US/09/912,976
CURRENT FILING DATE: 2001-07-05
PRIOR APPLICATION NUMBER: 60/221,336
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/238,333
PRIOR FILING DATE: 2000-10-05
PRIOR APPLICATION NUMBER: 60/260,675
PRIOR FILING DATE: 2001-01-10
PRIOR APPLICATION NUMBER: 60/271,025
PRIOR FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: 60/278,164
PRIOR FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: 60/280,876
PRIOR FILING DATE: 2001-04-02
NUMBER OF SEQ ID NOS: 99
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 64
LENGTH: 316
TYPE: PRT
ORGANISM: Homo sapiens
US-09-912-976-64

Query Match 1.6%; Score 8; DB 12; Length 316;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
Db 274 LFYAVGTP 281

RESULT 11
US-09-965-422-45
Sequence 45, Application US/09965422
Publication No. US20030216545A1
GENERAL INFORMATION:

```
APPLICANT: Spytek, Kimberly A
APPLICANT: Casman, Stacie
APPLICANT: Padigaru, Muselidhara
APPLICANT: Dickson, Kevin
APPLICANT: Vernet, Corine
APPLICANT: Spaderna, Steven K
APPLICANT: Shenoy, Suresh G
APPLICANT: Gerlach, Valerie
APPLICANT: Ellerman, Karen
APPLICANT: Edinger, Shlomit
APPLICANT: MacDougall, John R
APPLICANT: Smithson, Glenda
APPLICANT: Li, Li
APPLICANT: Majumdar, Urial M
APPLICANT: Taylor, Sarah
APPLICANT: Gunther, Erik
APPLICANT: Tchermeyev, Velizar T
TITLE OF INVENTION: No. US20030216545A1 Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21401-132
CURRENT APPLICATION NUMBER: US/09/965,422
CURRENT FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/236,286
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: 60/236,284
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: 60/237,581
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/238,735
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/240,736
PRIOR FILING DATE: 2000-10-16
PRIOR APPLICATION NUMBER: 60/260,019
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/260,338
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 60/262,156
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 60/262,498
PRIOR FILING DATE: 2001-01-18
PRIOR APPLICATION NUMBER: 60/263,133
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/263,691
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/266,109
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/271,634
PRIOR FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 127
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 45
LENGTH: 316
TYPE: PRT
ORGANISM: Homo sapiens
US-09-965-422-45

Query Match 1.6%; Score 8; DB 12; Length 316;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
FILE OF INVENTION: MODULATORS OF BONE HOMEOSTASIS IDENTIFIED IN A HIGH-THROUGHPUT S.C.
FILE REFERENCE: 25,294-B USA
CURRENT APPLICATION NUMBER: US/10/225,630
CURRENT FILING DATE: 2002-08-22
PRIOR APPLICATION NUMBER: US 60/314,056
PRIOR FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: US 60/356,935
PRIOR FILING DATE: 2002-02-14
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 666
TYPE: PRT
ORGANISM: human
US-10-225-630-10

Query Match 1.6%; Score 8; DB 15; Length 666;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

390 GTTPIMLA 397
Db 364 GTTPIMLA 371

RESULT 13
US-10-032-482-23
Sequence 23, Application US/10032482
Publication No. US20020197270A1
GENERAL INFORMATION:
APPLICANT: Rotter, Irvin
APPLICANT: Cohen, Yada
APPLICANT: Molkowicz, Roland
APPLICANT: Ruiz, Pedro
APPLICANT: Erez-Alon, Neta
APPLICANT: Herxler, Johannes
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS FOR INDUCTION OF ANTI-TUMOR
FILE REFERENCE: COHEN42
CURRENT APPLICATION NUMBER: US/10/032,482
CURRENT FILING DATE: 2002-01-02
PRIOR APPLICATION NUMBER: US/09/445,602
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: PCT/IL98/00266
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: IL 121041
PRIOR FILING DATE: 1997-06-09
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn version 3.0
SEQ ID NO 23
LENGTH: 20
TYPE: PRT
ORGANISM: Mus musculus
US-10-032-482-23

Query Match 1.4%; Score 7; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

431 QKVGKA 437
Db 14 QKVGKA 20

RESULT 14
US-09-864-761-36243
Sequence 36243, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
```


TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

PRIOR FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263, 6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

SEQ ID NO 36243

LENGTH: 48

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AC012665.1

OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2

OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8

OTHER INFORMATION: EST HUMAN HIT: BE535927.1, EVALUATE 5.10e+00

OTHER INFORMATION: SWISSPROT HIT: P40418, EVALUATE 6.00e+00

US-09-864-761-36243

Query Match

Best Local Similarity 1.4%; Score 7; DB 9; Length 48;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 287 ILTLTL 293

Db 35 ILTLTL 41

RESULT 15

US-09-764-853-614

Sequence 614, Application US/09764853

Patent No. US20020090672A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: P206

CURRENT APPLICATION NUMBER: US/09/764,853

PRIOR FILING DATE: 2001-01-17

PRIOR APPLICATION data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 939

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 614

LENGTH: 59

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (5)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (59)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-853-614

Query Match

Best Local Similarity 1.4%; Score 7; DB 9; Length 59;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 136 GLGLGVA 142

Db 7 GLGLGVA 13

RESULT 16

US-09-764-898-271

Sequence 271, Application US/09764898

Patent No. US20020090673A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: P201

CURRENT APPLICATION NUMBER: US/09/764,898

PRIOR FILING DATE: 2001-01-17

PRIOR APPLICATION data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 311

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 271

LENGTH: 59

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (5)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (59)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-898-271

Query Match

Best Local Similarity 1.4%; Score 7; DB 9; Length 59;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 136 GLGLGVA 142

Db 7 GLGLGVA 13

RESULT 17

US-10-073-865-100

Sequence 100, Application US/10073865

Publication No. US20030044904A1

GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: P209C1
CURRENT APPLICATION NUMBER: US/10/073,865
CURRENT FILING DATE: 2002-02-14
Prior Application removed - See file Wrapper or Palm
NUMBER OF SEQ ID NOS: 154
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 100
LENGTH: 59
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (5)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
LOCATION: (59)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-073-865-100

Query Match 1.4%; Score 7; DB 15; Length 59;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 GLIGLVA 142
|||
Db 7 GLIGLVA 13

RESULT 18
US-09-393-634-11
Sequence 11, Application US/09393634
Patent No. US20020051997A1
GENERAL INFORMATION:
APPLICANT: Zuker, Charles S.
APPLICANT: Adler, Jon Elliot
APPLICANT: Ryba, Nick
APPLICANT: Mueller, Ken
APPLICANT: Hoon, Mark
APPLICANT: The Regents of the University of California
APPLICANT: The Government of the United States of America
APPLICANT: as represented by the Secretary of the
APPLICANT: Department of Health and Human Services
TITLE OF INVENTION: SF, a No. US2002005197A1el Family of Taste Receptors
FILE REFERENCE: 02307E-098000US
CURRENT APPLICATION NUMBER: US/09/393,634
CURRENT FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 77
TYPE: PRT
ORGANISM: Mus sp.
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (1)..(77)
OTHER INFORMATION: Xaa = any amino acid
US-09-393-634-11

Query Match 1.4%; Score 7; DB 9; Length 77;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 LTLTLVI 295
|||
Db 37 LTLTLVI 43

RESULT 19
US-10-383-982-11

Sequence 11, Application US/10383982
Publication No. US20030157568A1
GENERAL INFORMATION:
APPLICANT: Zuker, Charles S.
APPLICANT: Adler, Jon Elliot
APPLICANT: Ryba, Nick
APPLICANT: Mueller, Ken
APPLICANT: Hoon, Mark
APPLICANT: The Regents of the University of California
APPLICANT: The Government of the United States of America
APPLICANT: as represented by the Secretary of the
APPLICANT: Department of Health and Human Services
TITLE OF INVENTION: SF, a No. US20030157568A1el Family of Taste Receptors
FILE REFERENCE: 02307E-098000US
CURRENT APPLICATION NUMBER: US/10/383,982
CURRENT FILING DATE: 2003-03-07
PRIOR APPLICATION NUMBER: US/09/393,634
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 77
TYPE: PRT
ORGANISM: Mus sp.
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (1)..(77)
OTHER INFORMATION: mouse GR01
OTHER INFORMATION: Xaa = any amino acid
US-10-383-982-11

Query Match 1.4%; Score 7; DB 12; Length 77;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 LTLTLVI 295
|||
Db 37 LTLTLVI 43

RESULT 20
US-10-144-929-125
Sequence 125, Application US/10144929
Publication No. US20030069405A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: 70 Human Secreted Proteins
FILE REFERENCE: P2014P1
CURRENT APPLICATION NUMBER: US/10/144,929
CURRENT FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: US/09/251,329
PRIOR FILING DATE: 1999-02-17
PRIOR APPLICATION NUMBER: PCT/US98/17044
PRIOR FILING DATE: 1998-08-18
NUMBER OF SEQ ID NOS: 257
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 125
LENGTH: 98
TYPE: PRT
ORGANISM: Homo sapiens
US-10-144-929-125

Query Match 1.4%; Score 7; DB 15; Length 98;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 TPLMLAV 398
|||
Db 47 TPLMLAV 53

RESULT 21
US-10-032-482-3

Sequence 3, Application US/10032482
Publication No. US20020197270A1
GENERAL INFORMATION:
APPLICANT: Cohen, Irun
APPLICANT: ROTTER, Varda
APPLICANT: Wolkowicz, Roland
APPLICANT: Ruiz, Pedro
APPLICANT: EREZ-ALON, Neta
APPLICANT: HERKEL, Johannes
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS FOR INDUCTION OF ANTI-TUMOR
FILE REFERENCE: COHEN42
CURRENT APPLICATION NUMBER: US/10/032,482
CURRENT FILING DATE: 2002-01-02
PRIOR APPLICATION NUMBER: US/09/445,602
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: PCT/IL98/00266
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: IL 121041
PRIOR FILING DATE: 1997-06-09
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 106
TYPE: PRT
ORGANISM: Mus musculus
US-10-032-482-3

Query Match 1.4%; Score 7; DB 14; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 431 QKVKGA 437
DB 53 QKVKGA 59

RESULT 22
US-09-991-936-38
Sequence 38, Application US/09991936
Publication No. US20030073827A1
GENERAL INFORMATION:
APPLICANT: Brandt, Kevin S.
APPLICANT: Gaines, Patrick J.
APPLICANT: Stinchcomb, Dan T.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
FILE REFERENCE: FC-6-C1
CURRENT APPLICATION NUMBER: US/09/991,936
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: US/09/543,668
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: 60/128,704
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 1959
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38
LENGTH: 135
TYPE: PRT
ORGANISM: Ctenocephalides felis
US-09-991-936-38

Query Match 1.4%; Score 7; DB 11; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 ELKVLTP 29
DB 64 ELKVLTP 70

RESULT 23

US-09-991-936-41
Sequence 41, Application US/09991936
Publication No. US20030073827A1
GENERAL INFORMATION:
APPLICANT: Brandt, Kevin S.
APPLICANT: Gaines, Patrick J.
APPLICANT: Stinchcomb, Dan T.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
FILE REFERENCE: FC-6-C1
CURRENT APPLICATION NUMBER: US/09/991,936
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: US/09/543,668
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: 60/128,704
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 1959
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 41
LENGTH: 135
TYPE: PRT
ORGANISM: Ctenocephalides felis
US-09-991-936-41

Query Match 1.4%; Score 7; DB 11; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 ELKVLTP 29
DB 64 ELKVLTP 70

RESULT 24
US-09-921-397-106
Sequence 106, Application US/09921397
Patent No. US20020151484A1
GENERAL INFORMATION:
APPLICANT: HYBRIGENICS
TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
TITLE OF INVENTION: applications thereof
FILE REFERENCE: B4809A, J42
CURRENT APPLICATION NUMBER: US/09/921,397
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: EP 00402225.7
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 156
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 106
LENGTH: 137
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-921-397-106

Query Match 1.4%; Score 7; DB 10; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 356 TGAIVTP 362
DB 3 TGAIVTP 9

RESULT 25
US-10-002-631C-6
Sequence 6, Application US/10002631C
Publication No. US20030157486A1
GENERAL INFORMATION:
APPLICANT: Graff, Jonathon M.
APPLICANT: Muenster, Matthew
TITLE OF INVENTION: METHODS TO IDENTIFY SIGNAL SEQUENCES

FILE REFERENCE: A34943 090495.0243
CURRENT APPLICATION NUMBER: US/10/002.631C
PRIOR FILING DATE: 2001-10-31
PRIOR APPLICATION NUMBER: 60/300,309
NUMBER OF SEQ ID NOS: 324
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 6
LENGTH: 144
TYPE: PRT
ORGANISM: Homo sapiens
US-10-002-631C-6

Query Match 1.4%; Score 7; DB 12; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 LLLTLVL 294
Db 64 LLLTLVL 70

RESULT 26
US-10-017-161-700
Sequence 700, Application US/10017161
Publication No. US20030143668A1
GENERAL INFORMATION:
APPLICANT: SUMA, MAKIKO
APPLICANT: ASAI, KIYOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
FILE REFERENCE: 084335/0152
CURRENT APPLICATION NUMBER: US/10/017.161
CURRENT FILING DATE: 2002-12-18
PRIOR APPLICATION NUMBER: JP 2001/246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2430
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 700
LENGTH: 174
TYPE: PRT
ORGANISM: Homo sapiens
US-10-017-161-700

Query Match 1.4%; Score 7; DB 12; Length 174;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 LLLTLVL 295
Db 46 LLLTLVL 52

RESULT 27
US-09-925-297-778
Sequence 778, Application US/09925297
Patent No. US20020081659A1
GENERAL INFORMATION:
APPLICANT: ROSEN ET AL.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA105
CURRENT APPLICATION NUMBER: US/09/925.297
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05989
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 928
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 778
LENGTH: 181
TYPE: PRT

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (145)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (155)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (163)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (169)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-778

Query Match 1.4%; Score 7; DB 9; Length 181;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLLASG 200
Db 112 ISLLASG 118

RESULT 28
US-09-738-626-6137
Sequence 6137, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738.626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377464
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 6137
LENGTH: 202
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-6137

Query Match 1.4%; Score 7; DB 10; Length 202;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GVVSVLI 340
Db 38 GVVSVLI 44

RESULT 29
US-10-153-668-424
Sequence 424, Application US/10153668
Publication No. US20030092616A1
GENERAL INFORMATION:
APPLICANT: HONDA, GOICHI

APPLICANT: MATSUDA, AKIO
APPLICANT: MURAMATSU, Shuji
APPLICANT: ISHIZAMA, Kenya
TITLE OF INVENTION: STATE Activating Gene
FILE REFERENCE: 1254-0207P
CURRENT APPLICATION NUMBER: US/10/153,668
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 60/293,172
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/316,031
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,403
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: JP 2001-157043
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: JP 2001-260681
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: JP 2001-313175
PRIOR FILING DATE: 2001-10-10
NUMBER OF SEQ ID NOS: 488
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 424
LENGTH: 209
TYPE: PRT
ORGANISM: Homo sapiens
US-10-153-668-424

Query Match 1.4%; Score 7; DB 15; Length 209;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 194 ISLASG 200
Db 49 ISLASG 55

RESULT 30
US-10-153-668-478
Sequence 478, Application US/10153668
Publication No. US20030092616A1
GENERAL INFORMATION:
APPLICANT: HONDA, Goichi
APPLICANT: MATSUDA, AKIO
APPLICANT: MURAMATSU, Shuji
APPLICANT: ISHIZAMA, Kenya
TITLE OF INVENTION: STATE Activating Gene
FILE REFERENCE: 1254-0207P
CURRENT APPLICATION NUMBER: US/10/153,668
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 60/293,172
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/316,031
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,403
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: JP 2001-157043
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: JP 2001-260681
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: JP 2001-313175
PRIOR FILING DATE: 2001-10-10
NUMBER OF SEQ ID NOS: 488
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 478
LENGTH: 209
TYPE: PRT
ORGANISM: Homo sapiens
US-10-153-668-478

Query Match 1.4%; Score 7; DB 15; Length 209;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 194 ISLASG 200
Db 49 ISLASG 55

RESULT 31
US-10-238-075-643
Sequence 643, Application US/10238075
Publication No. US20030146324A1
GENERAL INFORMATION:
APPLICANT: I.N.S.E.R.M.
TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolate
FILE REFERENCE: BLANDINE
CURRENT APPLICATION NUMBER: US/10/238,075
CURRENT FILING DATE: 2002-09-10
PRIOR APPLICATION NUMBER: 0003145
PRIOR FILING DATE: 2000-03-10
NUMBER OF SEQ ID NOS: 1576
SOFTWARE: PatentIn version 3.1
SEQ ID NO 643
LENGTH: 234
TYPE: PRT
ORGANISM: Escherichia coli
US-10-238-075-643

Query Match 1.4%; Score 7; DB 12; Length 234;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 139 GLVAILR 145
Db 36 GLVAILR 42

RESULT 32
US-10-153-668-314
Sequence 314, Application US/10153668
Publication No. US20030092616A1
GENERAL INFORMATION:
APPLICANT: HONDA, Goichi
APPLICANT: MATSUDA, AKIO
APPLICANT: MURAMATSU, Shuji
APPLICANT: ISHIZAMA, Kenya
TITLE OF INVENTION: STATE Activating Gene
FILE REFERENCE: 1254-0207P
CURRENT APPLICATION NUMBER: US/10/153,668
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 60/293,172
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/316,031
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,403
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: JP 2001-157043
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: JP 2001-260681
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: JP 2001-313175
PRIOR FILING DATE: 2001-10-10
NUMBER OF SEQ ID NOS: 488
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 314
LENGTH: 253
TYPE: PRT
ORGANISM: Homo sapiens
US-10-153-668-314

Query Match 1.4%; Score 7; DB 15; Length 253;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 194 ISLASG 200

Db 49 ISLASG 55

RESULT 33
US-10-153-668-476
; Sequence 476, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STATE Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153, 668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293, 172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316, 031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328, 403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 476
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-476

Query Match 1.4%; Score 7; DB 15; Length 253;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLASG 200
Db 49 ISLASG 55

RESULT 34
US-10-156-136-35
; Sequence 35, Application US/10156136
; Publication No. US20030128696A1
; GENERAL INFORMATION:
; APPLICANT: NI, et al.
; TITLE OF INVENTION: Polynucleotides and Polypeptides Encoding
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Ave
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/156,136
; FILING DATE: 29-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,146

FILING DATE: <Unknown>
APPLICATION NUMBER: WO US98/00959
FILING DATE: 21-JAN-1998
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF354PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-10-156-136-35

Query Match 1.4%; Score 7; DB 16; Length 253;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLASG 200
Db 49 ISLASG 55

RESULT 35
US-09-925-301-1234
; Sequence 1234, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1234
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1234

Query Match 1.4%; Score 7; DB 9; Length 282;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLASG 200
Db 78 ISLASG 84

RESULT 36
US-10-151-336-3
; Sequence 3, Application US/10151336
; Publication No. US20030079248A1
; GENERAL INFORMATION:
; APPLICANT: Mason, Hugh

APPLICANT: Palmer, Kenneth
APPLICANT: Hefferon, Kathleen
APPLICANT: Mor, Tsafir
APPLICANT: Antzen, Charles
TITLE OF INVENTION: Gemini Virus Vectors for Gene Expression in Plants
FILE REFERENCE: 4868/84453
CURRENT APPLICATION NUMBER: US/10/151,336
CURRENT FILING DATE: 2002-05-20
PRIOR APPLICATION NUMBER: US/09/414,276
PRIOR FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 293
TYPE: PRT
ORGANISM: bean yellow dwarf virus
US-10-151-336-3

Query Match 1.4%; Score 7; DB 15; Length 293;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 TPELIF 106
Db 39 TPELIF 45

RESULT 37
US-09-835-788A-19
Sequence 19, Application US/09835788A
Patent No. US2002007458A1
GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: Death Domain-Containing Receptor Polynucleotides, Polypeptides, a
FILE REFERENCE: PTO18P1
CURRENT APPLICATION NUMBER: US/09/835,788A
CURRENT FILING DATE: 2001-04-17
PRIOR APPLICATION NUMBER: PCT/US00/28666
PRIOR FILING DATE: 2000-10-17
PRIOR APPLICATION NUMBER: 60/159,585
PRIOR FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: 60/167,246
PRIOR FILING DATE: 1999-11-24
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 19
LENGTH: 306
TYPE: PRT
ORGANISM: Homo sapiens
US-09-835-788A-19

Query Match 1.4%; Score 7; DB 9; Length 306;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 TPELAV 398
Db 190 TPELAV 196

RESULT 38
US-10-175-042-19
Sequence 19, Application US/10175042
Publication No. US20030181710A1
GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: Death Domain-Containing Receptor Polynucleotides,
TITLE OF INVENTION: Polypeptides, and
FILE REFERENCE: PTO18P1
CURRENT APPLICATION NUMBER: US/10/175,042
CURRENT FILING DATE: 2002-06-20

PRIOR APPLICATION NUMBER: 09/835,788
PRIOR FILING DATE: 2001-04-17
PRIOR APPLICATION NUMBER: PCT/US00/28666
PRIOR FILING DATE: 2000-10-17
PRIOR APPLICATION NUMBER: 60/159,585
PRIOR FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: 60/167,246
PRIOR FILING DATE: 1999-11-24
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 19
LENGTH: 306
TYPE: PRT
ORGANISM: Homo sapiens
US-10-175-042-19

Query Match 1.4%; Score 7; DB 12; Length 306;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 TPELAV 398
Db 190 TPELAV 196

RESULT 39
US-10-146-772-12
Sequence 12, Application US/10146772
Publication No. US20030124698A1
GENERAL INFORMATION:
APPLICANT: Weiner, David
APPLICANT: Chaplin, Jennifer
APPLICANT: Chi, Ellen
APPLICANT: Milan, Aileen
APPLICANT: Desantis, Grace
APPLICANT: Madden, Mark
APPLICANT: Burk, Mark
TITLE OF INVENTION: Nitric Oxide
FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
CURRENT APPLICATION NUMBER: US/10/146,772
CURRENT FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: US 60/309,006
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US 60/351,336
PRIOR FILING DATE: 2002-01-22
PRIOR APPLICATION NUMBER: US 60/300,189
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 09/751,299
PRIOR FILING DATE: 2000-12-28
PRIOR APPLICATION NUMBER: US 60/254,414
PRIOR FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/173,609
PRIOR FILING DATE: 1999-12-29
NUMBER OF SEQ ID NOS: 386
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 321
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-12

Query Match 1.4%; Score 7; DB 15; Length 321;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 KDTLIV 54
Db 7 KDTLIV 13

RESULT 40
US-10-208-018-104
; Sequence 104, Application US/10208018
; Publication No. US20030115632A1
; GENERAL INFORMATION:
; APPLICANT: Lardizabal, Kathryn D
; APPLICANT: Hawkins, Deborah J
; APPLICANT: Thompson, Gregory A
; TITLE OF INVENTION: Diacylglycerol Acyltransferase Proteins
; FILE REFERENCE: 16515.155
; CURRENT APPLICATION NUMBER: US/10/208.018
; CURRENT FILING DATE: 2002-07-31
; PRIOR APPLICATION NUMBER: US 10/121,857
; PRIOR FILING DATE: 2002-04-15
; PRIOR APPLICATION NUMBER: US 09/345,461
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,631
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/130,829
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 104
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Glycine max
US-10-208-018-104

Query Match 1.4%; Score 7; DB 15; Length 329;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 194 ISLASG 200
Db 178 ISLASG 184

Search completed: November 25, 2003, 10:13:27
Job time : 57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 06:54:49 ; Search time 53 Seconds

(without alignments)
1542.343 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 515

Sequence: 1 MKTKBKPRGKLRSLFMPH.....AQSALKEGVAGQDSAPASS 515

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched:

1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : A Geneseq 190un03.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description |
|------------|-------|-------------|-----------|----|-------------|
| 1 | 515 | 100.0 | 515 | 21 | AAV90265 |
| 2 | 313 | 60.8 | 524 | 20 | AAV34951 |
| 3 | 29 | 5.6 | 529 | 20 | AAV37153 |
| 4 | 11 | 2.1 | 523 | 21 | AAV31017 |
| 5 | 11 | 2.1 | 567 | 21 | AAV50990 |
| 6 | 11 | 2.1 | 618 | 21 | AAV31016 |
| 7 | 11 | 2.1 | 655 | 21 | AAV31015 |
| 8 | 10 | 1.9 | 569 | 21 | AAV50991 |
| 9 | 9 | 1.7 | 498 | 21 | AAV50992 |

| | | | | | | |
|----|---|-----|------|----|----------|---------------------|
| 10 | 8 | 1.6 | 58 | 22 | AAO06244 | Human polypeptide |
| 11 | 8 | 1.6 | 137 | 20 | AAV32859 | Human heat shock p |
| 12 | 8 | 1.6 | 138 | 22 | ABG13197 | Novel human diagno |
| 13 | 8 | 1.6 | 138 | 22 | AAU30541 | Novel human secret |
| 14 | 8 | 1.6 | 143 | 24 | ABP76567 | Human GENSER prote |
| 15 | 8 | 1.6 | 147 | 23 | ABP41536 | Human ovarian anti |
| 16 | 8 | 1.6 | 173 | 21 | AAV51067 | Murine TCRbeta pro |
| 17 | 8 | 1.6 | 197 | 18 | AAW20458 | H. pylori transmem |
| 18 | 8 | 1.6 | 198 | 18 | AAW20930 | H. pylori surface |
| 19 | 8 | 1.6 | 207 | 23 | AAE27063 | Mouse TCR Jintubet |
| 20 | 8 | 1.6 | 222 | 22 | AAU33158 | Novel human secret |
| 21 | 8 | 1.6 | 247 | 17 | AAE85467 | PHLV-L env protein |
| 22 | 8 | 1.6 | 293 | 7 | AAE60238 | Sequence of the be |
| 23 | 8 | 1.6 | 293 | 10 | AAW91815 | Mammalian T lympho |
| 24 | 8 | 1.6 | 293 | 18 | AAW01532 | Cytotoxic T lympho |
| 25 | 8 | 1.6 | 293 | 21 | AAV50114 | Murine cytotoxic T |
| 26 | 8 | 1.6 | 302 | 23 | AAE25850 | Murine TCR B prote |
| 27 | 8 | 1.6 | 303 | 20 | AAV05403 | Killer T-cell rece |
| 28 | 8 | 1.6 | 306 | 23 | ABG73647 | Murine beta-TCR mu |
| 29 | 8 | 1.6 | 312 | 22 | AAV71456 | Human olfactory re |
| 30 | 8 | 1.6 | 314 | 20 | AAV05728 | Mouse A1 T cell re |
| 31 | 8 | 1.6 | 316 | 22 | AAV71665 | Human olfactory re |
| 32 | 8 | 1.6 | 316 | 22 | AAV72012 | Human olfactory re |
| 33 | 8 | 1.6 | 316 | 22 | AAV72199 | Human olfactory re |
| 34 | 8 | 1.6 | 316 | 22 | AAV72973 | Olfactory receptor |
| 35 | 8 | 1.6 | 316 | 23 | ABP95755 | Human GPCR polypep |
| 36 | 8 | 1.6 | 316 | 23 | ABT04015 | Human G-protein co |
| 37 | 8 | 1.6 | 316 | 23 | ABT04016 | Human G-protein co |
| 38 | 8 | 1.6 | 317 | 21 | AAV72904 | Human olfactory re |
| 39 | 8 | 1.6 | 318 | 22 | AAV72904 | Human olfactory re |
| 40 | 8 | 1.6 | 320 | 22 | AAV72622 | Murine OR-like pol |
| 41 | 8 | 1.6 | 320 | 22 | AAV72905 | Murine OR-like pol |
| 42 | 8 | 1.6 | 326 | 22 | AAV72626 | Murine OR-like pol |
| 43 | 8 | 1.6 | 343 | 23 | AAV71457 | Human olfactory re |
| 44 | 8 | 1.6 | 343 | 22 | AAV72371 | Human OR-like pol |
| 45 | 8 | 1.6 | 347 | 22 | AAV72625 | Murine OR-like pol |
| 46 | 8 | 1.6 | 348 | 22 | AAV72624 | Murine OR-like pol |
| 47 | 8 | 1.6 | 351 | 22 | AAV72621 | Murine OR-like pol |
| 48 | 8 | 1.6 | 356 | 20 | AAV78160 | Human secreted pro |
| 49 | 8 | 1.6 | 356 | 22 | AAV92659 | Human protein sequ |
| 50 | 8 | 1.6 | 358 | 22 | AAV78566 | Human protein sequ |
| 51 | 8 | 1.6 | 360 | 21 | AAV58321 | Arabidopsis thalia |
| 52 | 8 | 1.6 | 378 | 21 | AAV58320 | Arabidopsis thalia |
| 53 | 8 | 1.6 | 378 | 22 | ABH11981 | Human Slp-2 homolo |
| 54 | 8 | 1.6 | 378 | 22 | AAV79550 | Human protein SEQ |
| 55 | 8 | 1.6 | 392 | 22 | ABG12913 | Novel human diagno |
| 56 | 8 | 1.6 | 400 | 10 | AAV93189 | Novel human diagno |
| 57 | 8 | 1.6 | 400 | 22 | AAV29332 | Novel mar regulate |
| 58 | 8 | 1.6 | 430 | 23 | ABG73650 | Murine single chai |
| 59 | 8 | 1.6 | 433 | 22 | ABG12912 | Novel human diagno |
| 60 | 8 | 1.6 | 444 | 22 | ABG13721 | Novel human diagno |
| 61 | 8 | 1.6 | 464 | 22 | AAU28241 | Novel human secret |
| 62 | 8 | 1.6 | 467 | 23 | ABP69216 | Human polypeptide |
| 63 | 8 | 1.6 | 485 | 18 | AAW23245 | Human angiotensino |
| 64 | 8 | 1.6 | 485 | 18 | AAU31120 | Novel human secret |
| 65 | 8 | 1.6 | 485 | 22 | AAV67350 | Human angiotensino |
| 66 | 8 | 1.6 | 485 | 22 | AAV48944 | Human angiotensino |
| 67 | 8 | 1.6 | 485 | 22 | AAV48945 | Human angiotensino |
| 68 | 8 | 1.6 | 485 | 22 | AAV48947 | Human angiotensino |
| 69 | 8 | 1.6 | 485 | 22 | AAV48948 | Human angiotensino |
| 70 | 8 | 1.6 | 485 | 22 | AAV48949 | Human angiotensino |
| 71 | 8 | 1.6 | 492 | 16 | AAV85471 | PTLV-L env glycopr |
| 72 | 8 | 1.6 | 666 | 21 | ABR39453 | Notch-intracellular |
| 73 | 8 | 1.6 | 1013 | 21 | AAV84687 | Amino acid sequenc |
| 74 | 8 | 1.6 | 1013 | 22 | AAE11053 | Asbhya gossypii GT |
| 75 | 8 | 1.6 | 1360 | 22 | ABE61753 | Drosophila melanog |
| 76 | 8 | 1.6 | 1664 | 20 | AAV95557 | Mus musculus notch |
| 77 | 8 | 1.6 | 3067 | 22 | AAU28053 | Novel human secret |
| 78 | 8 | 1.4 | 16 | 15 | AAV45246 | HCV NS5 region (24 |
| 79 | 8 | 1.4 | 16 | 15 | AAV45247 | HCV NS5 region (24 |
| 80 | 8 | 1.4 | 16 | 20 | AAV10136 | T cell epitope/MHC |
| 81 | 7 | 1.4 | 16 | 23 | ABG79819 | MHC class I molecu |
| 82 | 7 | 1.4 | 20 | 20 | AAV89167 | Anti-p53 monoclonal |

| | | | | | | |
|-----|---|-----|----|----|----------|--------------------|
| 83 | 7 | 1.4 | 21 | 23 | ABB84104 | Naja naja atra shc |
| 84 | 7 | 1.4 | 25 | 23 | AAM51808 | HCV protease inhib |
| 85 | 7 | 1.4 | 30 | 23 | AAU84759 | HCV Hepc1a segment |
| 86 | 7 | 1.4 | 37 | 22 | AAO09368 | Human polypeptide |
| 87 | 7 | 1.4 | 48 | 22 | ABG50377 | Human liver peptid |
| 88 | 7 | 1.4 | 48 | 22 | ABR30346 | Peptide #2997 enco |
| 89 | 7 | 1.4 | 48 | 22 | ABR35518 | Peptide #3024 enco |
| 90 | 7 | 1.4 | 48 | 22 | ABR20945 | Protein #2944 enco |
| 91 | 7 | 1.4 | 48 | 22 | AAM56335 | Human brain expres |
| 92 | 7 | 1.4 | 48 | 22 | AAM68709 | Human bone marrow |
| 93 | 7 | 1.4 | 48 | 22 | AAM16537 | Peptide #2971 enco |
| 94 | 7 | 1.4 | 48 | 22 | AAM29017 | Peptide #3054 enco |
| 95 | 7 | 1.4 | 48 | 22 | AAM04253 | Peptide #2935 enco |
| 96 | 7 | 1.4 | 48 | 23 | ABG38292 | Human peptid enco |
| 97 | 7 | 1.4 | 52 | 21 | ABR45031 | Human secreted pro |
| 98 | 7 | 1.4 | 57 | 23 | ABP03855 | Human ORFX protein |
| 99 | 7 | 1.4 | 59 | 22 | ABR10306 | Human cDNA SEQ ID |
| 100 | 7 | 1.4 | 59 | 22 | AAU18124 | Novel human uterin |

ALIGNMENTS

RESULT 1

AAV90265 standard; Protein; 515 AA.

ID AAV90265 standard; Protein; 515 AA.

AAV90265;

22-SEP-2000 (first entry)

C. pneumoniae ATP/ADP translocase protein sequence.

ATP/ADP translocase; Chlamydia infection; diagnosis; therapy.

Chlamydia pneumoniae.

W0200039157-A1.

06-JUL-2000.

22-DEC-1999; 99WO-CA01224.

28-DEC-1998; 98US-0114060.

12-MAR-1999; 99US-0123967.

30-JUN-1999; 99US-0141271.

(CONN-) CONNAUGHT LAB LTD.

Murdin AD, Oomen RP, Wang J, Dunn P;

WPI; 2000-452368/39.

N-PSDB; AAA30922.

Novel Chlamydia polynucleotides and polypeptides, useful for diagnosis,

prevention and treatment of Chlamydia infection in mammals -

Claim 16; Fig 1; 81pp; English.

This sequence represents the Chlamydia pneumoniae ATP/ADP translocase

the DNA or protein, are useful for diagnosing, preventing or treating

Chlamydia infection. The sequences can also be used in a method for

the detection of Chlamydia infection. Primers or probes derived from the

DNA sequence are useful in diagnostic tests for detecting Chlamydia

infection.

Sequence 515 AA;

| | | | |
|----|-----|--|-----|
| QY | 1 | MTETKEKPEKGLRSFLMPHTHETLKKVLPMFLMPCITENYTVLRDPTKDTLIVGAPSSGA | 60 |
| DB | 1 | MTETKEKPEKGLRSFLMPHTHETLKKVLPMFLMPCITENYTVLRDPTKDTLIVGAPSSGA | 60 |
| QY | 61 | EALPFIKFLVPCALIFMLIVAKLSNLSKQALFYAVGTPFLIFPALPFTVIYPLRDVL | 120 |
| DB | 61 | EALPFIKFLVPCALIFMLIVAKLSNLSKQALFYAVGTPFLIFPALPFTVIYPLRDVL | 120 |
| QY | 121 | HPTEPADRLQALIPGLGLVAILRWTFPAFVYLAELMGSVNLSLMFGFANEITKHE | 180 |
| DB | 121 | HPTEPADRLQALIPGLGLVAILRWTFPAFVYLAELMGSVNLSLMFGFANEITKHE | 180 |
| QY | 181 | AKRFYALFGIGANISLASGRAIWMASKLRASVSEGVDPMGISRLMAMTIVSGVLMA | 240 |
| DB | 181 | AKRFYALFGIGANISLASGRAIWMASKLRASVSEGVDPMGISRLMAMTIVSGVLMA | 240 |
| QY | 241 | SYMWNKNVLTDRPFYNPEEMQKKGAKPKAMKDSFLYLDSPYILLLTLVIAVIGIC | 300 |
| DB | 241 | SYMWNKNVLTDRPFYNPEEMQKKGAKPKAMKDSFLYLDSPYILLLTLVIAVIGIC | 300 |
| QY | 301 | INLEVTWKSQTLQYPMNDYSEPMGNFSFMTGVSVLIMLFVGVNIRKFGMLGALV | 360 |
| DB | 301 | INLEVTWKSQTLQYPMNDYSEPMGNFSFMTGVSVLIMLFVGVNIRKFGMLGALV | 360 |
| QY | 361 | TPVMVLLTGIVFPFALVIFRNQASGLVAMFGTTPMLAVVGAIONLSKSTKALPDSTK | 420 |
| DB | 361 | TPVMVLLTGIVFPFALVIFRNQASGLVAMFGTTPMLAVVGAIONLSKSTKALPDSTK | 420 |
| QY | 421 | EMAYIPLDQEQKVKRAIDVVAARFGKSGGALIOGGLVIGCSIGAMPYLAIVILFI | 480 |
| DB | 421 | EMAYIPLDQEQKVKRAIDVVAARFGKSGGALIOGGLVIGCSIGAMPYLAIVILFI | 480 |
| QY | 481 | AIWLVSATKLNKFLAQSALKEQVVAQEDSAPASS | 515 |
| DB | 481 | AIWLVSATKLNKFLAQSALKEQVVAQEDSAPASS | 515 |

RESULT 2

AAV34951 standard; Protein; 524 AA.

ID AAV34951 standard; Protein; 524 AA.

AAV34951;

13-SEP-1999 (first entry)

Chlamydia pneumoniae transport polypeptide.

Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

vaccine; neutralising epitope.

Chlamydia pneumoniae.

W09927105-A2.

03-JUN-1999.

20-NOV-1998; 98WO-1B01890.

04-NOV-1998; 98US-0107078.

21-NOV-1997; 97FR-0014673.

(BEST) GENSET.

Griffais R;

WPI; 1999-357842/30.

Genome sequence of Chlamydia pneumoniae

Page 871-872; Disclosure; 1912pp; English.

AAV34584-Y35879 represent the proteins encoded by all the open reading

CC frames in the complete genome (see AAX31990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.

XX SQ Sequence 524 AA;

Query Match Best Local Similarity 99.6%; Score 313; DB 20; Length 524;

Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTKTEKPKGKLSFLMPTHTHEKLVLMFMEFCITFNTYVLRDTKTLIVGAPGSGA 60
 DB 10 MTKTEKPKGKLSFLMPTHTHEKLVLMFMEFCITFNTYVLRDTKTLIVGAPGSGA 69
 QY 61 EALPFIKFLVNCATIFMLIYAKLSILSKALFYAVGPPLIFPALFPTVIYPLRDVL 120
 DB 70 EALPFIKFLVNCATIFMLIYAKLSILSKALFYAVGPPLIFPALFPTVIYPLRDVL 129
 QY 121 HPTFADRLQAIIPGGLGVAILRMTFAFYVLAELMGSWMLSLFMGFANEITKHE 180
 DB 130 HPTFADRLQAIIPGGLGVAILRMTFAFYVLAELMGSWMLSLFMGFANEITKHE 189
 QY 181 AKGFYALFGIGANISLASGRAIVWASKLRASVEGVPWGISRLMANTIVSGVLMA 240
 DB 190 AKGFYALFGIGANISLASGRAIVWASKLRASVEGVPWGISRLMANTIVSGVLMA 249
 QY 241 SYWINKNVLTDRFYNPEMOKGKAKRPMNMKOSFLYLDSPYLLTLVIAYGIC 300
 DB 250 SYWINKNVLTDRFYNPEMOKGKAKRPMNMKOSFLYLDSPYLLTLVIAYGIC 309
 QY 301 INIETVWKSQKLQYRNMDYSEFMGNFSFMTGVSVLIMLFVGNVIRKFGMLTGALV 360
 DB 310 INIETVWKSQKLQYRNMDYSEFMGNFSFMTGVSVLIMLFVGNVIRKFGMLTGALV 369
 QY 361 TPWVVLITGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKYALPDSTK 420
 DB 370 TPWVVLITGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKYALPDSTK 429
 QY 421 EMAYIPLDQOKYKKAIDVVAARFGKSGALIQGSLVTCGSIAMTPYLAIVILFIT 480
 DB 430 EMAYIPLDQOKYKKAIDVVAARFGKSGALIQGSLVTCGSIAMTPYLAIVILFIT 489
 QY 481 AITWLSATKLNKLFIAQSALKEQVNAQEDSAPASS 515
 DB 490 AITWLSATKLNKLFIAQSALKEQVNAQEDSAPASS 524

RESULT 3

ID AAY37153 standard; Protein; 529 AA.

XX AC AAY37153;

DT 07-OCT-1999 (first entry)

XX CHlamydia trachomatis transport protein.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KM paratrachoma; inclusion conjunctivitis; genital disease; peritrophic;
 KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KM bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX Chlamydia trachomatis.

XX MO9928475-A2.

XX

PD 10-JUN-1999.

XX 27-NOV-1998; 98MO-IB01939.

XX 04-NOV-1998; 98US-0107077.

PR 28-NOV-1997; 97PR-0015041.

PR 17-DEC-1997; 97PR-0016034.

XX (GSET) GENSET.

XX Griffais R;

XX MPI; 1999-371125/31.

PT Genome sequence of Chlamydia trachomatis

XX Disclosure; Page 935-936; 1755pp; English.

XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome

CC of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as

CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences

CC can also be used to control growth of the microorganism. Chlamydia

CC trachomatis is responsible for a large number of diseases, e.g. eye

CC diseases such as conventional trachoma, nonendemic trachoma,

CC paratrachoma, and inclusion conjunctivitis; genital diseases such as

CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,

CC peritrophic, bartholinitis; pneumopathy in breast feeding infants;

CC and venereal lymphogranulomatosis. The polypeptides of the invention

CC may be of use in treating these diseases.

XX Sequence 529 AA;

Query Match Best Local Similarity 100.0%; Score 29; DB 20; Length 529;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 AYIPLDQOKYKKAIDVVAARFGKSGG 451

DB 422 AYIPLDQOKYKKAIDVVAARFGKSGG 450

RESULT 4

ID AAG13017 standard; Protein; 523 AA.

XX AAG13017;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 12353.

KM Protein identification; signal transduction pathway; metabolic pathway;

KM hybridisation assay; genetic mapping; gene expression control; promoter;

KM termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135622.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 10-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148585.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151087.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152263.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155569.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158832.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.1%; Score 11; DB 21; Length 523;

Best Local Similarity 100.0%; Pred. No. 0.07; 0; Indels 0; Gaps 0;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 KVKGAIDVV 442
414 KVKGAIDVV 424

Db 414 KVKGAIDVV 424

RESULT 5
AA50990
ID AAY50990 standard; Protein; 587 AA.

XX AAY50990;

DT 10-MAR-2000 (first entry)

DE A. thaliana ATP1 protein.

KM ADP/ATP translocator; transgenic plant; plasmid; oil; starch;

KW yield increase; ATP1.

OS Arabidopsis thaliana.

PN WO958654-A2.

PD 18-NOV-1999.

PF 12-MAY-1999; 99WO-EP03292.

PR 13-MAY-1998; 98DB-1021442.

XX (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.

PA (PLAC) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.

PI Neuhaus E, Moehlmann T, Graeve-Kampfenkel K, Tjaden J, Schell J;
PI Martini N;

DR WPI; 2000-062295/05.

XX Transgenic plants with modified plasmid ADP-ATP translocator activity

PS Example 2; Fig 3; 60pp; German.

CC This invention describes a novel genetically modified transgenic plant
CC cell, where the genetic modification is the introduction of a foreign
CC nucleic acid molecule, which exists, or its expression leads, to increase

CC of plasmid ADP/ATP translocation activity in particular compared to wild
CC type plant cells. Nucleic acid molecules encoding a plasmid ADP/ATP
CC translocator can be used to create transgenic plants and plant
CC cells with increased yields, especially an increased oil and/or starch
CC content. The plants are able to synthesize a modified starch. Plants,
CC such as maize, with an increased oil and/or starch content are of
CC commercial interest especially for the wet milling industry. Increased
CC starch content is important in food crops and in the paper and textile
CC industries. This sequence represents the Arabidopsis thaliana ATP1
CC protein which is used in the method of the invention.

XX Sequence 587 AA;

Query Match 2.1%; Score 11; DB 21; Length 587;

Best Local Similarity 100.0%; Pred. No. 0.078; 0; Indels 0; Gaps 0;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 KVKGAIDVV 442
509 KVKGAIDVV 519

Db 509 KVKGAIDVV 519

RESULT 6
AAG13016
ID AAG13016 standard; Protein; 618 AA.

XX AAG13016;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 12352.

KM Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142350.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150684.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158322.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0160814.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

[illegible]

| | | |
|----|-------------|--------------|
| PR | 04-JUN-1999 | 99US-0137528 |
| PR | 04-JUN-1999 | 99US-0137528 |
| PR | 07-JUN-1999 | 99US-0137724 |
| PR | 08-JUN-1999 | 99US-0138094 |
| PR | 10-JUN-1999 | 99US-0138540 |
| PR | 10-JUN-1999 | 99US-0138617 |
| PR | 14-JUN-1999 | 99US-0139119 |
| PR | 16-JUN-1999 | 99US-0139452 |
| PR | 17-JUN-1999 | 99US-0139453 |
| PR | 17-JUN-1999 | 99US-0139453 |
| PR | 18-JUN-1999 | 99US-0139454 |
| PR | 18-JUN-1999 | 99US-0139455 |
| PR | 18-JUN-1999 | 99US-0139456 |
| PR | 18-JUN-1999 | 99US-0139456 |
| PR | 18-JUN-1999 | 99US-0139463 |
| PR | 18-JUN-1999 | 99US-0139750 |
| PR | 18-JUN-1999 | 99US-0139763 |
| PR | 21-JUN-1999 | 99US-0139817 |
| PR | 22-JUN-1999 | 99US-0139817 |
| PR | 23-JUN-1999 | 99US-0140353 |
| PR | 23-JUN-1999 | 99US-0140354 |
| PR | 24-JUN-1999 | 99US-0140695 |
| PR | 28-JUN-1999 | 99US-0140923 |
| PR | 29-JUN-1999 | 99US-0140981 |
| PR | 30-JUN-1999 | 99US-0141287 |
| PR | 01-JUL-1999 | 99US-0141842 |
| PR | 01-JUL-1999 | 99US-0142154 |
| PR | 02-JUL-1999 | 99US-0142055 |
| PR | 06-JUL-1999 | 99US-0142390 |
| PR | 08-JUL-1999 | 99US-0142803 |
| PR | 09-JUL-1999 | 99US-0142920 |
| PR | 12-JUL-1999 | 99US-0142977 |
| PR | 13-JUL-1999 | 99US-0143354 |
| PR | 14-JUL-1999 | 99US-0143624 |
| PR | 15-JUL-1999 | 99US-0144005 |
| PR | 16-JUL-1999 | 99US-0144085 |
| PR | 16-JUL-1999 | 99US-0144085 |
| PR | 19-JUL-1999 | 99US-0144325 |
| PR | 19-JUL-1999 | 99US-0144331 |
| PR | 19-JUL-1999 | 99US-0144332 |
| PR | 19-JUL-1999 | 99US-0144333 |
| PR | 19-JUL-1999 | 99US-0144335 |
| PR | 20-JUL-1999 | 99US-0144352 |
| PR | 20-JUL-1999 | 99US-0144632 |
| PR | 20-JUL-1999 | 99US-0144884 |
| PR | 21-JUL-1999 | 99US-0144814 |
| PR | 21-JUL-1999 | 99US-0145086 |
| PR | 21-JUL-1999 | 99US-0145088 |
| PR | 22-JUL-1999 | 99US-0145085 |
| PR | 22-JUL-1999 | 99US-0145089 |
| PR | 22-JUL-1999 | 99US-0145091 |
| PR | 23-JUL-1999 | 99US-0145145 |
| PR | 23-JUL-1999 | 99US-0145218 |
| PR | 23-JUL-1999 | 99US-0145224 |
| PR | 26-JUL-1999 | 99US-0145276 |
| PR | 27-JUL-1999 | 99US-0145513 |
| PR | 27-JUL-1999 | 99US-0145518 |
| PR | 27-JUL-1999 | 99US-0145519 |
| PR | 28-JUL-1999 | 99US-0145551 |
| PR | 02-AUG-1999 | 99US-0146386 |
| PR | 02-AUG-1999 | 99US-0146388 |
| PR | 03-AUG-1999 | 99US-0147038 |
| PR | 04-AUG-1999 | 99US-0147204 |
| PR | 04-AUG-1999 | 99US-0147302 |
| PR | 05-AUG-1999 | 99US-0147192 |

```

PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147410.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149923.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155653.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160960.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 2.1%; Score 11; DB 21; Length 655;

```

Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 KVKGKALIDV 442
Db 546 KVKGKALIDV 556

RESULT 8
AAV50991 standard; Protein; 569 AA.
ID AAV50991;
AC AAV50991;
XX 10-MAR-2000 (first entry)
DT 10-MAR-2000 (first entry)
DE A. thaliana ATP2 protein.
XX
XX ADP/ATP translocator; transgenic plant; plastid; oil; starch;
KM yield increase; ATP2.
XX
OS Arabidopsis thaliana.
XX
XX WO9958654-A2.
XX
XX 18-NOV-1999.
PD 18-NOV-1999.
PF 12-MAY-1999; 99WO-EP03292.
XX
XX 13-MAY-1999; 98DE-1021442.
PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX Neuhaus E, Moehlmann T, Graeve-Kampfenkel K, Tjaden J, Schell J,
PI Martini N;
XX WPI; 2000-062235/05.
XX
XX Transgenic plants with modified plastid ADP-ATP translocator activity
PT
XX
XX Example 2; Fig 3; 60bp; German.
XX
XX This invention describes a novel genetically modified transgenic plant
CC cell, where the genetic modification is the introduction of a foreign
CC nucleic acid molecule, which exists, or its expression leads, to increase
CC of plastid ADP/ATP translocation activity in particular compared to wild
CC type plant cells. Nucleic acid molecules encoding a plastid ADP/ATP
CC translocator can be used to create transgenic plants and plant
CC cells with increased yields, especially an increased oil and/or starch
CC content. The plants are able to synthesize a modified starch. Plants,
CC such as maize, with an increased oil and/or starch content are of
CC commercial interest especially for the wet milling industry. Increased
CC starch content is important in food crops and in the paper and textile
CC industries. This sequence represents the Arabidopsis thaliana ATP2
CC protein which is used in the method of the invention.
XX
XX Sequence 569 AA;
SQ

Query Match 1.9%; Score 10; DB 21; Length 569;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 GKSGLIQQ 456
Db 522 GKSGLIQQ 531

RESULT 9
AAV50992 standard; Protein; 498 AA.
ID AAV50992
XX

```


AC AAY5092; (first entry)
 XX 10-MAR-2000
 DT R. prowazekii ATP homologue TLC protein.
 DE ADP/ATP translocator; transgenic plant; plastid; oil; starch;
 KM yield increase; TLC protein.
 XX Rickettsia prowazekii.
 OS WO958654-A2.
 PN 18-NOV-1999.
 PD 12-MAY-1999; 99WO-EF03292.
 PF 13-MAY-1998; 98DE-1021442.
 PR (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA Neuhaus E, Moehmann T, Graewe-Kampfenkel K, Tjaden J, Schell J;
 PI Martini N;
 XX MPI; 2000-062295/05.
 DR Transgenic plants with modified plastid ADP-ATP translocator activity
 XX -
 PT Example 2; Fig 3; 60pp; German.
 PS This invention describes a novel genetically modified transgenic plant
 CC cell, where the genetic modification is the introduction of a foreign
 CC nucleic acid molecule, which exists, or its expression leads, to increase
 CC of plastid ADP/ATP translocation activity in particular compared to wild
 CC type plant cells. Nucleic acid molecules encoding a plastid ADP/ATP
 CC translocator can be used to create transgenic plants and plant
 CC cells with increased yields, especially an increased oil and/or starch
 CC content. The plants are able to synthesize a modified starch. Plants,
 CC such as maize, with an increased oil and/or starch content are of
 CC commercial interest especially for the wet milling industry. Increased
 CC starch content is important in food crops and in the paper and textile
 CC industries. This sequence represents the Rickettsia prowazekii TLC
 CC protein which is used in the method of the invention.
 CC
 SQ Sequence 498 AA;
 Query Match 1.7%; Score 9; DB 21; Length 498;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 246 NKNVLTDP 254
 DB 241 NKNVLTDP 249
 RESULT 10
 AAO6244
 ID AAO6244 standard; Protein; 58 AA.
 XX
 AC AAO6244;
 XX
 DT 06-NOV-2001 (first entry)
 DE Human polypeptide SEQ ID NO 20136.
 XX
 KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;
 KM nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.

XX
 PN WO200164835-A2.
 XX 07-SEP-2001.
 PD 26-FEB-2001; 2001WO-US04927.
 PF 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Drmanac RT;
 PI MPI; 2001-514838/56.
 XX N-PSDB; AA186175.
 DR Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX Claim 20; SEQ ID NO 20136; 1399pp + Sequence Listing; English.
 PS The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 58 AA;
 Query Match 1.6%; Score 8; DB 22; Length 58;
 Best Local Similarity 100.0%; Pred. No. 8.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 128 RLQALILP 135
 DB 25 RLQALILP 32
 RESULT 11
 AAY32859
 ID AAY32859 standard; Protein; 137 AA.
 XX
 AC AAY32859;
 XX
 DT 01-NOV-1999 (first entry)
 DE Human heat shock protein homologue 2 (HSPH-2) protein sequence.
 XX
 KM Human heat shock protein homologue; HSPH-1; HSPH-2; diagnosis; cancer;
 KM inflammation; AIDS; Crohn's disease; allergy; therapy; drug screening.
 XX
 OS Homo sapiens.
 XX
 PN US5945287-A.
 PD 31-AUG-1999.
 PF 19-NOV-1997; 97US-0974546.
 PR 19-NOV-1997; 97US-0974546.
 XX (INCY-) INCYTE PHARM INC.

XX
PI Corley NC, Hillman JL, Lal P, Shah P;
XX
DR WPI, 1999-526253/44.
DR N-PSDB; AAZ11067.
XX
PT New molecular chaperones useful for diagnosis, prevention and
PT treatment of cancer and inflammation
XX
PS Claim 8; Fig 3; 35pp; English.
XX
CC This sequence is the human heat shock protein homologue, HSPH-2 of
CC the invention. The HSPH-1 polynucleotide was isolated from the BRATUT21
CC cDNA library constructed from cancerous brain tissue. The HSPH-2 DNA
CC sequence was isolated from the BRSTNOT18 cDNA library constructed from
CC diseased breast tissue. The HSPH polynucleotides and their complements
CC are useful for diagnosis of conditions or disorders associated with HSPH
CC expression. HSPH complements are also useful as antagonists for
CC prevention or treatment of cancer e.g. adenocarcinoma, breast and skin,
CC and inflammation associated with e.g. AIDS, Crohn's disease and
CC allergies. HSPH-1 and HSPH-2 are useful for drug screening using
CC libraries of compounds. The HSPH-1, HSPH-2 nucleotide sequences, and
CC their complements, are useful for detecting upstream sequences
CC e.g. promoters and regulatory elements. Vectors containing the DNA
CC sequences are useful for the preparation of HSPH proteins and mRNA
CC probes.
XX
SQ Sequence 137 AA;
Query Match 1.6%; Score 8; DB 20; Length 137;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 195 SLASGRA 202
Db 36 SLASGRA 43
RESULT 12
ABG13197
ID ABG13197 standard; Protein; 138 AA.
XX
AC ABG13197;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #1318.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS77364.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX
PS Claim 20; SEQ ID No 43556; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 138 AA;
Query Match 1.6%; Score 8; DB 22; Length 138;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 443 AARFGKSG 450
Db 56 AARFGKSG 63
RESULT 13
AAU30541
ID AAU30541 standard; Protein; 138 AA.
XX
AC AAU30541;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #1032.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
PR 26-JAN-2001; 2001US-0770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
PS Claim 20; Page 307; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The

CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
SQ Sequence 138 AA;
Query Match 1.6%; Score 8; DB 22; Length 138;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 195 SLIAGSRA 202
DB 33 SLIAGSRA 40
|||||
RESULT 14
ABP76367
ID ABP76367 standard; Protein; 143 AA.
XX
AC ABP76367;
XX
XX 21-FEB-2003 (first entry)
XX
XX Human GENSET protein SEQ ID 917.
XX
XX
XX Cytostatic; antiinflammatory; neurotropic; neuroprotective; cardiac;
XX gastrointestinal; gene therapy; GENSET; heavy metal toxicity; cancer;
XX inflammatory disease; immune disorder; neuromuscular; toxicity;
XX central nervous system; cardiovascular; gastrointestinal.
XX
XX Homo sapiens.
XX
XX WO200283898-A1.
XX
XX 24-OCT-2002.
XX
XX 18-APR-2001; 2001WO-IB00914.
XX
XX 18-APR-2001; 2001WO-IB00914.
XX
XX (GENSET) GENSET.
XX
XX Benjamin S, Tanaka H, Dumas Milne Edwards J, Jobert S, Giordano J;
XX WPI, 2003-075548/07.
XX
XX New GENSET polynucleotides and polypeptides, useful for treating heavy
XX metal toxicity, cancer, inflammatory diseases, immune disorders, and
XX the neuromuscular, CNS, cardiovascular or gastrointestinal effects of
XX the toxicity -
XX
XX Claim 14; Page 730; 735pp; English.
XX
XX The present invention relates to novel GENSET polynucleotides
XX (AB236404-AB236911) encoding polypeptides (ABP75963-ABP76368). The
XX polynucleotides and polypeptides are useful in screening and diagnostic
XX assays for abnormal GENSET expression and/or biological activity. They
XX are also useful for screening of compounds for treating or preventing
XX GENSET-related disorders, such as heavy metal toxicity, cancer,
XX inflammatory diseases, immune disorders, and the neuromuscular, central

CC nervous system (CNS), cardiovascular or gastrointestinal effects of the
CC toxicity.
XX
SQ Sequence 143 AA;
Query Match 1.6%; Score 8; DB 24; Length 143;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 195 SLIAGSRA 202
DB 36 SLIAGSRA 43
|||||
RESULT 15
ABP41536
ID ABP41536 standard; Protein; 147 AA.
XX
AC ABP41536;
XX
XX 22-AUG-2002 (first entry)
XX
XX Human ovarian antigen HSPSF84, SEQ ID NO:2668.
XX
XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
XX ovarian cancer; breast cancer; tumour; reproductive system disorder;
XX infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
XX PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
XX inflammatory condition; immune disorder; blood disorder;
XX cardiovascular disorder; respiratory disorder; neurological disorder;
XX gastrointestinal disorder; urinary system disorder; drug screening;
XX Gene therapy; chromosome mapping; forensic analysis;
XX antibody preparation; cytostatic; immunomodulatory; neuroprotective;
XX antiinflammatory; gynaecological; reproductive.
XX
XX Homo sapiens.
XX
XX WO200200677-A1.
XX
XX 03-JAN-2002.
XX
XX 07-JUN-2001; 2001WO-US18569.
XX
XX 07-JUN-2000; 2000US-209467P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Birse CR, Rosen CA;
XX
XX WPI, 2002-147878/19.
XX
XX N-PSDB; AB054613.
XX
XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
XX useful in the prevention, treatment and diagnosis of cancer (e.g.
XX ovarian cancer), immune disorders, cardiovascular disorders and
XX neurological diseases -
XX
XX Claim 11; SEQ ID No 2668; 2922pp; English.
XX
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
XX ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also
XX encompasses polypeptides 90% identical and polynucleotides 95% identical
XX to the sequences of the invention. The invention additionally relates to
XX recombinant vectors and host cells comprising human ovarian antigen
XX polynucleotides, antibodies against human ovarian antigens, and the use
XX of ovarian antigen polynucleotides and polypeptides in diagnosing,
XX treating, prognosing or preventing various ovary and/or breast-related
XX disorders. Such conditions include ovarian cancer and breast cancer, and
XX metastatic tumours of ovarian or breast origin, reproductive system
XX disorders (e.g., infertility, disorders of pregnancy, anovulation,
XX polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
XX disorders, infections (e.g., chlamydia, HIV, toxoplasmosis), and toxic
XX shock syndrome), inflammatory conditions (e.g., mastitis, coprophitis and

CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 147 AA;

Query Match 1.6%; Score 8; DB 23; Length 147;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 SLASGRA 202
 DB 46 SLASGRA 53

RESULT 16
 ID AAY51067 standard; protein; 173 AA.
 XX
 AC AAY51067;
 XX
 DT 22-MAR-2000 (first entry)
 XX
 DE Murine TCRbeta protein fragment #2.
 XX
 KM T cell receptor beta; TCRbeta; murine constant region;
 KM immunosuppressor.
 XX
 OS Mus sp.
 XX
 JF11302239-A.
 XX
 PD 02-NOV-1999.
 XX
 PF 21-APR-1998; 98JP-0110607.
 XX
 PR 21-APR-1998; 98JP-0110607.
 XX
 PA (KIRI) KIRIN BREWERY KK.
 XX
 DR WPI; 2000-075345/07.
 XX
 PT T cell receptor beta chain constant region peptide - has
 PT immunosuppressing activity
 XX
 PS Claim 2; Page 10-11; 15pp; Japanese.
 XX
 CC This invention describes a novel polypeptide comprising substantially
 CC part or all of the constant region of T cell receptor beta chain
 CC (TCRbeta) and containing substantially no other regions of TCRbeta and
 CC having immunosuppressing activity. The TCRbeta requires no consideration
 CC of tissue-compatible antigen for the patient to be dosed nor of antigen
 CC specificity. This sequence represents a fragment of the murine TCRbeta
 CC protein described in the method of the invention.

XX
 SQ Sequence 173 AA;

Query Match 1.6%; Score 8; DB 21; Length 173;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGLVMA 240
 DB 159 VSGLVMA 166

RESULT 17
 ID AAM20458 standard; Protein; 197 AA.
 XX
 AC AAM20458;
 XX
 DT 11-JUL-1997 (first entry)
 XX
 DE H. pylori transmembrane protein, 36134661.aa.
 XX
 KM Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KM identification; binding compound; bacterium; life cycle; activator;
 KM bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
 KM diagnosis.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9640893-A1.
 XX
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Mellgaerd BL,
 XX
 DR WPI; 1997-052306/05.
 DR N-PSDB; AAT67628.
 XX
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 73; Page 628; 1481pp; English.
 XX
 CC This sequence represents a H. pylori protein likely to contain three
 CC membrane spanning regions.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.

XX
 SQ Sequence 197 AA;

Query Match 1.6%; Score 8; DB 18; Length 197;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 YIILLTL 293
 DB 178 YIILLTL 185

RESULT 18
 AAM20930

ID AAW20930 standard; protein; 198 AA.
 AC AAW20930;
 DT 21-JUL-1997 (first entry)
 DE H. pylori surface or membrane protein, 16ep10117orf7.
 XX
 XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KM binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KM duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 XX
 XX Helicobacter pylori.
 OS
 XX WO9640893-A1.
 PN
 XX 19-DEC-1996.
 PD
 XX 06-JUN-1996; 96WO-US09122.
 PF
 XX 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 XX (ASTR) ASTRA AB.
 PA
 XX Berglindh OF, Smith D, Møllgaard BL;
 PI WPI, 1997-05306/05.
 DR N-PSDB; AAT68183.
 XX
 XX Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PT
 XX Claim 73; Page 1327; 1481pp; English.
 PS
 XX The present sequence is a H. pylori surface or membrane protein likely
 CC to contain at least three membrane spanning regions.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 CC
 XX
 SQ Sequence 198 AA;
 QY Query Match 1.6%; Score 8; DB 18; Length 198;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 286 YIIILLTLL 293
 179 YIIILLTLL 186
 RESULT 19
 AAE27063
 ID AAE27063 standard; Protein; 207 AA.
 XX
 XX AAE27063;
 AC
 XX 13-DEC-2002 (first entry)
 DT
 DE Mouse TCR Jintubeta2.6-C protein.
 XX

KM Mouse; T cell receptor; TCR; mesenchymal cell growth; cell therapy;
 KM carcinoma; wound healing; intronic J region; joining region; J region;
 KM constant domain; C domain; receptor; Jintubeta2.6-C.
 XX
 XX Mus sp.
 OS
 XX Key
 FH Region
 FT Location/Qualifiers
 FT 1..19
 FT /note= "Intron 5' to Ubeta2.6 region"
 FT 20..34
 FT /note= "Ubeta2.6 region"
 FT Region
 FT 35..38
 FT /note= "Cbeta2 region"
 XX
 XX WO20026636-A2.
 PN
 XX 29-AUG-2002.
 PD
 XX 20-FEB-2002; 2002WO-IL00130.
 PF
 XX 20-FEB-2001; 2001IL-0141539.
 PR
 XX (YEDA) YEDA RES & DEV CO LTD.
 PA (YBBA-) UNIV BAR-ILAN.
 XX
 XX Zilport D, Rozenzajn AL, Barda-Saad M, Shay-Tal Y;
 PI WPI, 2002-667067/71.
 DR N-PSDB; AAD44572.
 XX
 XX New isolated polynucleotide comprising a transcript of a T cell
 PT receptor gene, useful for transfecting mesenchymal cells for wound
 PT healing or suppression of carcinomas -
 PT
 XX Claim 18; Fig 1; 63pp; English.
 PS
 XX The invention relates to polynucleotide transcripts of T cell receptor
 CC (TCR) gene, as well as protein sequences encoded by these transcripts,
 CC and their use in the modulation of mesenchymal cell growth. The
 CC invention also concerns cDNA molecules encoded by a TCR gene lacking V
 CC region sequences and comprising a constant (C) domain and joining (J)
 CC region sequences and a 5' intronic J sequences upstream to the J region
 CC sequence including an in-frame methionine. Polynucleotides of the
 CC invention are useful for transfecting mesenchymal cells for wound
 CC healing or suppression of carcinomas. Antibodies of the invention are
 CC useful as markers of mesenchymal cells. The invention is useful in cell
 CC therapy. The present sequence is mouse TCR Jintubeta2.6-C protein.
 CC
 XX
 SQ Sequence 207 AA;
 QY Query Match 1.6%; Score 8; DB 23; Length 207;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 233 VSGIVLMA 240
 193 VSGIVLMA 200
 RESULT 20
 AAU33158
 ID AAU33158 standard; Protein; 222 AA.
 XX
 XX AAU33158;
 AC
 XX 18-DEC-2001 (first entry)
 DT
 DE Novel human secreted protein #3649.
 XX
 XX Human; vaccination; gene therapy; nutritional supplement;
 KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX

```

OS Homo sapiens.
XX
XX WO200179449-A2.
XX
XX 25-OCT-2001.
XX
XX 16-APR-2001; 2001WO-US08656.
XX
XX 18-APR-2000; 2000US-0552929.
XX
XX 26-JAN-2001; 2001US-0770160.
XX
XX (HYSEQ-) HYSEQ INC.
XX
XX Tang YF, Liu C, Demanac RT;
XX
XX WPI; 2001-611725/70.
XX
XX Nucleic acids encoding a range of human polypeptides, useful in genetic
XX vaccination, testing and therapy -
XX
XX Claim 20; Page 716; 765pp; English.
XX
XX The invention relates to novel human secreted polypeptides. The
XX polypeptides and antibodies to the polypeptides are useful for
XX determining the presence of or predisposition to a disease associated
XX with altered levels of polypeptide. The polypeptides are also useful for
XX identifying agents (agonists and antagonists) that bind to them. Cells
XX expressing the proteins are useful for identifying a therapeutic agent
XX for use in treatment of a pathology related to aberrant expression or
XX physiological interactions of the polypeptide. Vectors comprising
XX the nucleic acids encoding the polypeptides and cells genetically
XX engineered to express them are also useful for producing the proteins.
XX The proteins are useful in genetic vaccination, testing and
XX therapy, and can be used as nutritional supplements. They may be used to
XX increase stem cell proliferation; to regulate haematopoiesis; and in
XX bone, cartilage, tendon and/or nerve tissue growth or regeneration;
XX immune suppression and/or stimulation; as anti-inflammatory agents; and
XX in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid
XX sequences of novel human secreted proteins of the invention.
XX
XX Sequence 222 AA;
XX
XX Query Match 1.6%; Score 8; DB 22; Length 222;
XX Best Local Similarity 100.0%; Pred. No. 31;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 195 SLASGRA 202
XX |||||
XX 17 SLASGRA 24
XX
XX RESULT 21
XX AAR85467
XX ID AAR85467 standard; Protein; 247 AA.
XX
XX AAR85467;
XX
XX 12-MAR-1996 (first entry)
XX
XX PHLV-L env protein.
XX
XX PHLV-L, HTLV; baboon; vaccine; diagnostic; env protein;
XX PH69.
XX
XX T-lymphotropic primate virus.
XX
XX Key Location/Qualifiers
XX Cleavage-site 69..70
XX /note= "env glycoprotein proteolytic cleavage site"
XX Modified-site 161
XX /label= N-glycosylation_site
XX
XX MO9526405-A1.

```

```

XX
XX 05-OCT-1995.
XX
XX 23-MAR-1995; 95WO-BE00026.
XX
XX 08-JUN-1994; 94NL-0000932.
XX
XX 25-MAR-1994; 94EP-0200804.
XX
XX (REGA-) STICHTING REGA VZW.
XX
XX Deemyter JHMA, Goubau PFJ;
XX
XX WPI; 1995-351325/45.
XX
XX N-PSDB; AAT05712.
XX
XX New T-lymphotropic primate virus PTLV-L isolated from baboon(s) -
XX homologous to human viruses, also related polypeptide(s), nucleic
XX acid, antibodies etc., useful in diagnostic assays and vaccines
XX
XX Example 2; Fig 1, 42pp; English.
XX
XX A protein (AAR85467) homologous to the env glycoprotein precursor of
XX HTLV1 and HTLV2 is encoded by the env gene (see AAT05712) of
XX a novel T-lymphotropic primate virus, PTLV-L, isolated from
XX an Eritrean wild baboon (Papio hamadryas). Epitopes of this
XX virus can be used in vaccine prodn. or in the development of
XX diagnostic antibodies.
XX
XX Sequence 247 AA;
XX
XX Query Match 1.6%; Score 8; DB 16; Length 247;
XX Best Local Similarity 100.0%; Pred. No. 35;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 480 IAIWLVA 487
XX |||||
XX 73 IAIWLVA 80
XX
XX RESULT 22
XX AAP60238
XX ID AAP60238 standard; Protein; 293 AA.
XX
XX AAP60238;
XX
XX 25-MAR-2003 (updated)
XX
XX 19-AUG-1991 (first entry)
XX
XX Sequence of the beta subunit of a clonally diversified
XX integral membrane protein encoded by clone pHS11.
XX
XX T-cell clone specific antibody; cytotoxic T lymphocyte; immunoassay;
XX drug delivery.
XX
XX Mus musculus.
XX
XX Key Location/Qualifiers
XX Peptide 1..11
XX /label= signal
XX
XX Region 12..107
XX /label= variable
XX
XX Region 108..120
XX /label= joining
XX
XX Region 121..266
XX /label= constant
XX
XX Region 267..288
XX /label= transmembrane
XX
XX Region 289..293
XX /label= cytoplasmic
XX
XX Disulfide-bond 34..102
XX Disulfide-bond 151..212
XX Disulfide-bond 247
XX /note= "inter-chain"

```

```

XX  EP180878-A.
XX  14-MAY-1986.
XX  26-OCT-1985; 85EP-0113658.
XX  31-OCT-1984; 84US-0666988.
XX  13-JUN-1984; 84US-0620122.
XX  01-MAR-1984; 84US-0585333.
XX  22-OCT-1984; 84US-0663809.
XX  27-JUL-1989; 89US-035897.
XX  (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX  (STRD ) UNIV LELAND STANFORD JUNIOR.
XX  Saito H, Kranz DM, Eisen HN, Tonegawa S;
XX  WPI; 1986-126342/20.
XX  N-PSDB; AAN60167.
XX  New hetero-dimeric T-lymphocyte receptor and its sub-units -
XX  useful in prodn. of T-cell clone specific antibodies for
XX  immunosays, isolation procedures, specific delivery of bound
XX  drugs etc.
XX  Example; Fig 4b; 50pp; English.
XX  T-cell specific cDNA clones were isolated from an alloreactive CTL
XX  clone 2C of Balb/c mouse origin and specific for the D end of the
XX  Balb/C H-2 complex. A library of cDNA was constructed and screened
XX  with hybridisation probes. Two distinct classes of T-cell-specific
XX  cDNA clones whose genes are rearranged in cytotoxic T lymphocytes
XX  (CTLs) were identified (see AAN60166 and AAN60167). A heterodimeric
XX  T lymphocyte receptor comprising an alpha and a beta subunit is
XX  claimed.
XX  (Updated on 25-MAR-2003 to correct PA field.)
XX  Sequence 293 AA;

QY  233 VSGLVVMA 240
DB  279 VSGLVVMA 286

Query Match 1.6%; Score 8; DB 7; Length 293;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 23
AAP91815
ID AAP91815 standard; protein; 293 AA.
XX
XX  AAP91815;
AC  25-MAR-2003 (updated)
XX  22-MAY-1990 (first entry)
XX
XX  Mammalian T lymphocyte receptor beta subunit.
DE  T lymphocyte receptor; alpha subunit; cytotoxic T cell; lymphoma;
XX  chemotherapy; PHDS11.
XX  Mus.
XX
XX  Key Location/Qualifiers
XX  FT 12..107
XX  FT /label=variable domain
XX  FT 108..120
XX  FT /label=joining domain
XX  FT 121..266
XX  FT /label=constant domain
XX  FT 267..288

```

```

FT  Domain /label=transmembrane domain
FT  289..293
FT  /label=cytoplasmic domain
FT  Disulfide-bond 34..102
FT  Disulfide-bond 151..212
FT  Disulfide-bond 247..247
FT  /note="links alpha chain to beta chain"
XX
XX  US4874845-A.
XX
XX  17-OCT-1989.
XX
XX  13-JUN-1984; 84US-0620122.
XX
XX  13-JUN-1984; 84US-0620122.
XX
XX  (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX  Saito H, Kranz DM, Eisen HN, Tonegawa S;
XX  WPI; 1989-377931/51.
XX  N-PSDB; AAN93897.
XX  Mammalian T lymphocyte receptor sub-unit - with structure derived using
XX  cDNA clones derived from functional murine cytotoxic T lymphocyte clone.
XX  Disclosure; Page -; 15pp; English.
XX
XX  The sequence was deduced from the DNA insert in PHSD11 isolated from the
XX  alloreactive CTL clone 2C of BALB.B (mouse) origin.
XX  See also AAP91815.
XX  (Updated on 25-MAR-2003 to correct PA field.)
XX  Sequence 293 AA;

QY  233 VSGLVVMA 240
DB  279 VSGLVVMA 286

Query Match 1.6%; Score 8; DB 10; Length 293;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 24
AAW01532
ID AAW01532 standard; Protein; 293 AA.
XX
XX  AAW01532;
AC  25-MAR-2003 (updated)
XX  14-APR-1997 (first entry)
XX
XX  Cytotoxic T lymphocyte clone 2C beta subunit cDNA.
DE  Cytotoxic T lymphocyte; CTL; T cell receptor; T helper cell;
XX  PHDS11.
XX  Mus sp.
XX
XX  Key Location/Qualifiers
XX  FT 1..11
XX  FT /label=Sig_peptide
XX  FT 12..293
XX  FT /label=Mat_protein
XX  FT 12..107
XX  FT /label=V
XX  FT /note="variable region (amino acids 1-96 of the
XX  FT 108..120
XX  FT /label=J
XX  FT /note="joining region (amino acids 97-109 of
XX  FT the mature beta-subunit)"

```

| | | |
|-----------------------|---|------------------------|
| FT | Region | 121..247 /label= C |
| FT | /note= "constant region (amino acids 110 to approx. | |
| FT | 236 of mature beta-subunit)" | |
| FT | Region | 266..290 /label= TM |
| FT | /note= "transmembrane region (amino acids 255-279 | |
| FT | of mature beta-subunit)" | |
| FT | Region | 289..293 /label= CY |
| FT | /note= "cytoplasmic region (amino acids 278..284 | |
| FT | of mature beta-subunit)" | |
| FT | Disulfide-bond | 34..102 |
| FT | Disulfide-bond | 151..212 |
| FT | Disulfide-bond | 247 |
| FT | /note= "Cys-247 (Cys-236 of mature beta-subunit) | |
| FT | forms a disulphide bond with Cys-234 of | |
| FT | mature gamma subunit" | |
| FN | USS580961-A. | |
| PD | 03-DEC-1996. | |
| PX | 18-MAR-1994; | 94US-0210326. |
| PX | 13-JUN-1984; | 84US-0620122. |
| PR | 14-NOV-1988; | 88US-0271217. |
| PR | 08-AUG-1991; | 91US-0742539. |
| PR | 18-MAR-1994; | 94US-0210326. |
| PA | (MASI) MASSACHUSETTS INST TECHNOLOGY. | |
| PX | Eisen HN, Kranz DM, Saito H, Tonegawa S; | |
| PI | WPI; 1997-033602/03. | |
| DR | N-PDB; AAT46176. | |
| XX | T cell receptor gamma subunit proteins - useful for antibody prodn. | |
| XX | Disclosure; Fig 4A; 19pp; English. | |
| CC | Amino acid sequences for the beta subunit (AAW01532) and gamma | |
| CC | subunit (AAW01533) of a heterodimeric T cell receptor were deduced | |
| CC | from cDNA clones (AAT46176-77) isolated from alloreactive cytotoxic | |
| CC | lymphocyte (CTL) clone 2C, of BALB.B origin and specific for the D | |
| CC | end of the BALB/c H-2 complex (d haplotype). Genes corresponding | |
| CC | to these cDNAs are expressed and rearranged specifically in T cells | |
| CC | and show homologies to immunoglobulin variable and constant region | |
| CC | genes. The constant region of the beta chain corresponds to the | |
| CC | beta chain of helper T cells and can be used to produce antibodies | |
| CC | useful in the isolation and identification of CTLs and T helper | |
| CC | cells. Recombinant beta and gamma chains can be produced in host | |
| CC | cells and used to raise T cell receptor-specific antibodies useful | |
| CC | e.g. for antibody-bound drugs delivery. | |
| CC | (Updated on 25-MAR-2003 to correct PF field.) | |
| XX | Sequence | 293 AA; |
| SQ | | |
| Query Match | 1.6%; Score 8; DB 18; Length 293; | |
| Best Local Similarity | 100.0%; Pred. No. 41; | |
| Matches | 8; Conservative 0; Mismatches 0; Indels 0; Gaps | |
| OY | 233 VSGVLMA 240 | |
| Db | | |
| | 279 VSGVLMA 286 | |
| RESULT 25 | | |
| AAVS0114 | standard; Protein; 293 AA. | |
| XX | AAVS0114; | |
| XX | | |

| DT | 09-FEB-2000 | (first entry) |
|-----------------------|---|---|
| XX | Murine cytotoxic T-lymphocyte (CTL) receptor beta subunit. | |
| DE | Cytotoxic; T-lymphocyte; alloreactive; receptor; subunit; beta; | |
| KW | major histocompatibility complex; MHC; antigen; recognition; | |
| KW | immunoglobulin; variable domain; hypervariable region; constant domain; | |
| KW | antibody. | |
| XX | | |
| OS | Mus sp. | |
| XX | | |
| FH | Key | Location/Qualifiers |
| FT | Peptide | 1..11 |
| FT | Protein | /label=Signal_peptide |
| FT | | 12..293 |
| FT | Domain | /note="Mature CTL receptor beta subunit" |
| FT | | 12..107 |
| FT | Disulfide-bond | /note="Variable domain" |
| FT | Domain | 34..102 |
| FT | | 108..120 |
| FT | Domain | /note="Joining region" |
| FT | | 121..266 |
| FT | Disulfide-bond | /note="Constant region" |
| FT | Disulfide-bond | 151..212 |
| FT | | 247 |
| FT | Domain | /note="Disulphide bond with Cys 252 of gamma subunit" |
| FT | | 267..288 |
| FT | Domain | /note="Transmembrane domain" |
| FT | | 289..293 |
| FT | Domain | /note="Cytoplasmic domain" |
| XX | | |
| PN | US5977321-A. | |
| XX | | |
| PD | 02-NOV-1999. | |
| XX | | |
| PF | 08-AUG-1991; | 91US-0742539. |
| XX | | |
| PR | 13-JUN-1984; | 84US-0620120. |
| PR | 14-NOV-1988; | 88US-0271217. |
| XX | | |
| PA | (MAST) MASSACHUSETTS INST TECHNOLOGY. | |
| XX | | |
| PI | Tonegawa S, Eisen HN, Saito H, Kranz DM; | |
| XX | | |
| DR | WPI; 2000-012127/01. | |
| XX | N-ESDB; AAZ32662. | |
| XX | | |
| PT | A new antibody to a T-lymphocyte receptor gamma subunit | - |
| XX | | |
| PS | Example; Fig 4A; 14pp; English. | |
| XX | | |
| CC | This sequence represents the beta subunit of a murine alloreactive | |
| CC | cytotoxic T-lymphocyte. CTL receptors comprise 2 subunits, beta | |
| CC | and gamma (AAV5015) and are responsible for the recognition | |
| CC | of antigens presented on major histocompatibility complex | |
| CC | (MHC) molecules. The beta and gamma subunits each comprise two | |
| CC | extracellular immunoglobulin (Ig)-like domains (an N-terminal | |
| CC | variable domain and a more C-terminal constant domain, both stabilised | |
| CC | by intramolecular disulphide bonds), a transmembrane domain and a | |
| CC | cytoplasmic cytoplasmic domain. In addition, the two subunits are linked by | |
| CC | an intermolecular disulphide bond. The invention relates to a novel | |
| CC | antibody against the non-hypervariable region of a mammalian CTL | |
| CC | receptor gamma subunit, which is common to CTL receptor subunits of | |
| CC | different specificity. Such antibodies are useful for T-cell receptor | |
| CC | identification and isolation, and for delivering antibody-bound drugs | |
| CC | to a specific cell. | |
| XX | | |
| XX | | |
| SQ | Sequence 293 AA; | |
| Query March | 1.6%; Score 8; DB 21; Length 293; | |
| Best Local Similarity | 100.0%; Pred. No. 41; | |
| Matches 8; | Conservative 0; Mismatches 0; Indels 0; Gaps 0; | |

QY 233 VSGIIVLMA 240
 |||||
 DB 279 VSGIIVLMA 286

RESULT 26
 AAE25850
 ID AAE25850 standard; Protein: 302 AA.

AC AAE25850;

DT 15-NOV-2002 (first entry)

DE Murine TCR B protein.

KW Human leukocyte antigen; HLA; T cell receptor; cytotoxic T lymphocyte;
 TCR; CTL; adoptive immunotherapy; tumour; gene therapy; cyostatic;
 murine.

OS Mus musculus.

PN US2002064521-A1.

PD 30-MAY-2002.

PF 22-FEB-2001; 2001US-0789697.

PR 22-FEB-2000; 2000US-183752P.

PA (CITY) CITY OF HOPE.

PI Ellenhorn JDI, Diamond DJ;

DR WPI; 2002-589174/63.

DR N-PSDB; AAD42573.

PT Novel nucleic acid encoding A chain or B chain of hu p53-specific,
 human leukocyte antigen restricted mu T cell receptor, useful for
 producing CD8+ cytotoxic lymphocytes for adoptive tumor immunotherapy

PT -

PS Claim 1; Page 17-18; 29pp; English.

CC The present invention relates to novel nucleic acids encoding A chain or
 B chain of human p53-specific, human leukocyte antigen (HLA) restricted
 murine T cell receptor (TCR). Human CD8+ cytotoxic T lymphocytes (CTLs)
 transfected with sequences of the invention are useful in adoptive
 immunotherapy for treating an individual having a tumour that over-
 expresses p53. Sequences of the invention are useful in gene therapy.

CC The present sequence is murine TCR B protein.

CC Sequence 302 AA;

Query Match 1.6%; Score 8; DB 23; Length 302;
 Best Local Similarity 100.0%; Pred. No. 42;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 233 VSGIIVLMA 240
 |||||
 DB 288 VSGIIVLMA 295

RESULT 27

AAV05403
 ID AAV05403 standard; Protein: 303 AA.

AC AAV05403;

DT 02-JUL-1999 (first entry)

DE Killer T-cell receptor protein sequence.

KW Killer T-cell receptor; HIV; infection; HIV-1 IIIB; therapy.

XX Mus musculus.
 OS
 XX
 PN WO9916885-A1.
 XX
 PD 08-APR-1999.
 XX

PF 28-SEP-1998; 98WO-JP04345.

PR 26-SEP-1997; 97JP-0262536.

PA (KYOW) KYOWA HAKKO KOGYO KK.
 (SAIT/) SAITO T.
 (TAKA/) TAKAHASHI H.

PI Saito T, Takahashi H;

DR WPI; 1999-255096/21.

DR N-PSDB; AAX36391.

PT Killer T-cell receptor peptide specifically recognizing HIV-infected
 cells

PS Claim 7; Page 60-62; 75pp; Japanese.

CC This sequence is a killer T-cell receptor protein of the invention,
 which recognises and damages cells infected with human immunodeficiency
 virus (HIV), especially with HIV-1 IIIB. The receptor can be used as
 a constituent of compositions for the treatment of HIV infection.

CC Sequence 303 AA;

Query Match 1.6%; Score 8; DB 20; Length 303;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGIIVLMA 240
 |||||
 DB 289 VSGIIVLMA 296

RESULT 28

ABG73647
 ID ABG73647 standard; Protein: 306 AA.

AC ABG73647;

DT 27-MAR-2003 (first entry)

DE Murine beta-TCR muvbeta-muvbeta.

KW Murine; T cell receptor; TCR; hdm2; T cell response; alpha TCR; beta TCR;
 antigen-recognition sequence; ARS; fusion construct; cyostatic;

KW apoptotic; tumour; leukaemia; immunisation.

OS Mus musculus.

PN DE10109854-A1.

PD 12-SEP-2002.

PF 01-MAR-2001; 2001DE-1009854.

PR 01-MAR-2001; 2001DE-1009854.

PA (STAN/) STANISLAWSKI T.

PI Theobalt M, Voss H, Stanislawski T;

DR WPI; 2002-714556/78.

PT New polypeptide of a murine alpha/beta T-cell receptor, useful for
 treating tumors and leukemia, induces specific lysis or apoptosis of

PT cells expressing hdm2 protein -
 XX
 PS Claim 1; Page 19-20; 52pp; German.
 XX
 CC This invention describes a novel murine alphabeta T-cell receptor (TCR)
 CC that mediates a hdm2 protein-specific T cell response, a fusion protein
 CC (FP) that includes the TCR and nucleic acid encoding it, alpha or
 CC beta-chains of a TCR that include the antigen-recognition sequence (ARS)
 CC of an antibody specific for aa 81-88 of hdm2 (or its complex with
 CC HLA-A2-specific antibody) and a method for identifying hdm2-specific
 CC antigens. The TCR of the invention has cytostatic and apoptotic activity.
 CC The products of the invention are useful for treatment, prevention and
 CC diagnosis of hdm2-associated diseases, particularly tumours and
 CC leukemia, including use for passive or active immunisation. They can
 CC also be used to screen for therapeutic agents. This sequence represents
 CC the murine beta-TCR muvbeta-mucbeta described in the disclosure of
 CC the invention.
 CC
 SQ Sequence 306 AA;
 Query Match 1.6%; Score 8; DB 23; Length 306;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 233 VSGELVMA 240
 |||||
 DB 292 VSGELVMA 299
 RESULT 29
 ID AAG71456
 AC AAG71456 standard; Protein; 312 AA.
 XX
 AC AAG71456;
 XX
 DT 30-JUN-2001 (first entry)
 XX
 DE Human olfactory receptor polypeptide, SEQ ID NO: 1137.
 XX
 KM Human, olfactory receptor; OR; primary scent determination;
 KM secondary scent determination; polypeptide library; odour receptor;
 KM scent profile; scent fingerprint; scent representation.
 XX
 OS Homo sapiens.
 XX
 FN WO200127158-A2.
 XX
 PD 19-APR-2001.
 XX
 PF 06-OCT-2000; 2000WO-US27582.
 XX
 PR 08-OCT-1999; 99US-0158615.
 PR 24-FEB-2000; 2000US-0184809.
 XX
 PA (DIGI-) DIGISCENTS.
 PA (YEDA) YEDA RES & DEV CO LTD.
 XX
 PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
 XX
 DR WPI; 2001-290713/30.
 XX
 PT New polynucleotides which encode polypeptides involved in olfactory
 PT sensation for identifying olfactory agonists and antagonists -
 XX
 PS Claim 11; Page 658-659; 1857pp; English.
 XX
 CC The present sequence is an olfactory receptor which is encoded by
 CC one of a number of novel polynucleotides. The polynucleotides can be
 CC used in screening for olfactory agonists and antagonists. The methods
 CC allow for the determination of primary scents and the identification
 CC of the odour receptors used to detect these primary scents. The methods
 CC also enable determination of secondary scents and the identification of
 CC combinations of odour receptors that are involved in detecting such

CC secondary scents. This enables the construction of a scent representation
 CC (also called a scent fingerprint or scent profile), which may be used to
 CC re-create and edit scents. Libraries of olfactory receptors are useful
 CC for determining the interaction pattern of a composition with the
 CC receptors, and can be used for determining differences in the olfactory
 CC faculties of different individuals.
 CC
 SQ Sequence 312 AA;
 Query Match 1.6%; Score 8; DB 22; Length 312;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 94 LFYAVGTP 101
 |||||
 DB 274 LFYAVGTP 281
 RESULT 30
 ID AAY05728
 AC AAY05728 standard; Protein; 314 AA.
 XX
 AC AAY05728;
 XX
 DT 19-JUN-1999 (first entry)
 XX
 DE Mouse A1 T cell receptor beta chain.
 XX
 KM T cell receptor beta chain; mouse; transgenic animal;
 KM animal model; immunological tolerance; graft rejection;
 KM tissue grafting; neonatal intolerance; transplantation antigen;
 KM H-Y antigen.
 XX
 OS Mus musculus.
 XX
 FN WO9916867-A1.
 XX
 PD 08-APR-1999.
 XX
 PF 30-SEP-1998; 98WO-GB02965.
 XX
 PR 01-OCT-1997; 97GB-0020868.
 XX
 PA (ISIS-) ISIS INNOVATION LTD.
 XX
 PI Cobbold SP, Waldmann H, Zelenika D;
 XX
 DR WPI; 1999-255090/21.
 DR N-PSDB; AAX25358.
 XX
 PT Transgenic non-human mammal having only CD4 positive T cells
 PT specific for at least one transplantation antigen, useful for
 PT studying immunological tolerance
 XX
 PS Disclosure; Page 35-36; 41pp; English.
 XX
 CC The present sequence represents the T cell receptor (TCR) beta
 CC chain from the A1 CD4+ T cell clone isolated from CBA/Ca mice. The
 CC A1 clone recognises the minor histocompatibility antigen H-Y
 CC present in male, but absent in female, mice. cDNA (see AAX25358)
 CC encoding the A1 TCR alpha chain was obtained by RT-PCR (see
 CC also AAX25352-53), and was used in the construction of A1(M)
 CC transgenic mice. The invention relates to a transgenic animal
 CC model comprising TCR alpha and beta chains. It provides a
 CC genetically modified non-human mammal having a population of CD4
 CC positive T cells specific for one or a limited number of selected
 CC antigens, including at least transplantation antigen capable of
 CC rejecting a tissue transplant containing the transplantation
 CC antigen and, if applicable, the other selected antigens. The
 CC animal has TCR genes which encode a TCR specific for the
 CC transplantation antigen. The animal is useful for studying
 CC immunological tolerance, especially the mechanisms of tolerance to,
 CC and the rejection of, tissue grafts, and in pregnancy. The animals

CC are also useful for testing agents for biological activity in
 CC promoting or reducing immunological tolerance.

XX
 SQ Sequence 314 AA;

Query Match 1.6%; Score 8; DB 20; Length 314;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGLVVMA 240
 |||||
 Db 300 VSGLVVMA 307

RESULT 31

AAAG71665
 ID AAG71665 standard; Protein; 316 AA.

AC AAG71665;

DT 30-JUL-2001 (first entry)

DE Human olfactory receptor polypeptide, SEQ ID NO: 1346.

XX Human; olfactory receptor; OR; primary scent determination;

KW secondary scent determination; polypeptide library; odour receptor;

XX scent profile; scent fingerprint; scent representation.

OS Homo sapiens.

PN WO200127158-A2.

PD 19-APR-2001.

PF 06-OCT-2000; 2000WO-US27582.

PR 08-OCT-1999; 99US-0158615.

PR 24-FEB-2000; 2000US-0184809.

XX (DIGI-) DIGISCENTS.

PA (YEDA) YEDA RES & DEV CO LTD.

PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

XX WPI; 2001-290713/30.

DR New polynucleotides which encode polypeptides involved in olfactory

PT sensation for identifying olfactory agonists and antagonists -

XX Claim 11, Page 831-832; 1857pp; English.

XX The present sequence is an olfactory receptor which is encoded by

CC one of a number of novel polynucleotides. The polynucleotides can be

CC used in screening for olfactory agonists and antagonists. The methods

CC allow for the determination of primary scents and the identification

CC of the odour receptors used to detect these primary scents. The methods

CC also enable determination of secondary scents and the identification of

CC combinations of odour receptors that are involved in detecting such

CC secondary scents. This enables the construction of a scent representation

CC (also called a scent fingerprint or scent profile), which may be used to

CC re-create and edit scents. Libraries of olfactory receptors are useful

CC for determining the interaction pattern of a composition with the

CC receptors, and can be used for determining differences in the olfactory

CC faculties of different individuals.

XX Sequence 316 AA;

QY Query Match 1.6%; Score 8; DB 22; Length 316;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
 |||||

Db 274 LFYAVGTP 281

RESULT 32

AAAG72012
 ID AAG72012 standard; Protein; 316 AA.

AC AAG72012;

DT 30-JUL-2001 (first entry)

DE Human olfactory receptor polypeptide, SEQ ID NO: 1693.

XX Human; olfactory receptor; OR; primary scent determination;

KW secondary scent determination; polypeptide library; odour receptor;

XX scent profile; scent fingerprint; scent representation.

OS Homo sapiens.

PN WO200127158-A2.

PD 19-APR-2001.

PF 06-OCT-2000; 2000WO-US27582.

PR 08-OCT-1999; 99US-0158615.

PR 24-FEB-2000; 2000US-0184809.

XX (DIGI-) DIGISCENTS.

PA (YEDA) YEDA RES & DEV CO LTD.

PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

XX WPI; 2001-290713/30.

DR New polynucleotides which encode polypeptides involved in olfactory

PT sensation for identifying olfactory agonists and antagonists -

XX Claim 11, Page 1104-1105; 1857pp; English.

XX The present sequence is an olfactory receptor which is encoded by

CC one of a number of novel polynucleotides. The polynucleotides can be

CC used in screening for olfactory agonists and antagonists. The methods

CC allow for the determination of primary scents and the identification

CC of the odour receptors used to detect these primary scents. The methods

CC also enable determination of secondary scents and the identification of

CC combinations of odour receptors that are involved in detecting such

CC secondary scents. This enables the construction of a scent representation

CC (also called a scent fingerprint or scent profile), which may be used to

CC re-create and edit scents. Libraries of olfactory receptors are useful

CC for determining the interaction pattern of a composition with the

CC receptors, and can be used for determining differences in the olfactory

CC faculties of different individuals.

XX Sequence 316 AA;

QY Query Match 1.6%; Score 8; DB 22; Length 316;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
 |||||

DT 30-JUL-2001 (first entry)

DE Human olfactory receptor polypeptide, SEQ ID NO: 1880.
XX
XX Human; olfactory receptor; OR; primary scent determination;
KW secondary scent determination; polypeptide library; odour receptor;
KM scent profile; scent fingerprint; scent representation.
XX
OS Homo sapiens.
XX
FN WO200127158-A2.
XX
PD 19-APR-2001.
XX
PF 06-OCT-2000; 2000WO-US27582.
XX
PR 08-OCT-1999; 99US-0158615.
PR 24-FEB-2000; 2000US-0184809.
XX
XX (DIGI-) DIGISCENTS.
PA (YEDA) YEDA RES & DEV CO LTD.
XX
PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX
DR WPI; 2001-290713/30.
XX
PT New polynucleotides which encode polypeptides involved in olfactory
PT sensation for identifying olfactory agonists and antagonists -
XX
XX Claim 11; Page 1245-1246; 1857pp; English.
XX
XX The present sequence is an olfactory receptor which is encoded by
CC one of a number of novel polynucleotides. The polynucleotides can be
CC used in screening for olfactory agonists and antagonists. The methods
CC allow for the determination of primary scents and the identification
CC of the odour receptors used to detect these primary scents. The methods
CC also enable determination of secondary scents and the identification of
CC combinations of odour receptors that are involved in detecting such
CC secondary scents. This enables the construction of a scent representation
CC (also called a scent fingerprint or scent profile), which may be used to
CC re-create and edit scents. Libraries of olfactory receptors are useful
CC for determining the interaction pattern of a composition with the
CC receptors, and can be used for determining differences in the olfactory
CC faculties of different individuals.
XX
SQ Sequence 316 AA;
QY
Best Local Similarity 1.6%; Score 8; DB 22; Length 316;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 94 LFYAVGTP 101
274 LFYAVGTP 281
RESULT 34
AAG72973
ID AAG72973 standard; Protein; 316 AA.
XX
AC AAG72973;
XX
DT 30-JUL-2001 (first entry)
XX
DE Olfactory receptor-like polypeptide, SEQ ID NO: 2655.
XX
XX Human; olfactory receptor; OR; primary scent determination;
KW secondary scent determination; polypeptide library; odour receptor;
KM scent profile; scent fingerprint; scent representation.
XX
OS Unidentified.
XX
PN WO200127158-A2.
XX
PD 19-APR-2001.

XX
PF 06-OCT-2000; 2000WO-US27582.
XX
PR 08-OCT-1999; 99US-0158615.
PR 24-FEB-2000; 2000US-0184809.
XX
XX (DIGI-) DIGISCENTS.
PA (YEDA) YEDA RES & DEV CO LTD.
XX
PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX
DR WPI; 2001-290713/30.
XX
PT New polynucleotides which encode polypeptides involved in olfactory
PT sensation for identifying olfactory agonists and antagonists -
XX
XX Example 6; Page 1794-1795; 1857pp; English.
XX
XX The present sequence is an olfactory receptor polypeptide which was
CC used as a query sequence in a database search of olfactory receptor
CC (OR)-like sequences. The invention relates to isolated polynucleotides
CC encoding polypeptides involved in olfactory sensation. The
CC polynucleotides can be used in screening for olfactory agonists and
CC antagonists. The methods allow for the determination of primary scents
CC and the identification of the odour receptors used to detect these
CC primary scents. The methods also enable determination of secondary scents
CC and the identification of combinations of odour receptors that are
CC involved in detecting such secondary scents. This enables the
CC construction of a scent representation (also called a scent fingerprint
CC or scent profile), which may be used to re-create and edit scents.
CC Libraries of olfactory receptors are useful for determining the
CC interaction pattern of a composition with the receptors, and can be
CC used for determining differences in the olfactory faculties of different
CC individuals.
XX
SQ Sequence 316 AA;
QY
Best Local Similarity 1.6%; Score 8; DB 22; Length 316;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 94 LFYAVGTP 101
274 LFYAVGTP 281
RESULT 35
ABP95755
ID ABP95755 standard; Protein; 316 AA.
XX
AC ABP95755;
XX
DT 06-MAR-2003 (first entry)
XX
DE Human GPCR polypeptide SEQ ID NO 320.
XX
XX Human; GPCR; G protein coupled receptor; signal transduction; olfactory;
KW drug development; gustatory; taste; fragrance; receptor.
XX
OS Homo sapiens.
XX
PN WO200216548-A2.
XX
PD 28-FEB-2002.
XX
PF 30-JUL-2001; 2001WO-IB01446.
XX
PR 04-AUG-2000; 2000JP-0237818.
PR 13-FEB-2001; 2001JP-0034434.
XX
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
PA Haga T, Takeda S, Mitaku S;
XX
PI

| | |
|-----------|--|
| XX | WP1: 2002-304118/34. |
| DR | N-PSDB; ABZ43029. |
| XX | |
| PT | Database global search for G protein-coupled receptors, proteins and |
| PT | encoded genes for studying in vivo signal transduction mechanism and |
| PT | identifying targets for drug development |
| PS | |
| PS | Claim 10, SEQ ID NO 320, 97tp + Sequence Listing; Japanese. |
| XX | |
| CC | The invention relates to a method for screening G protein-coupled |
| CC | receptor (GPCR) genes (ABZ42870-ABZ43216) and/or GPCR proteins |
| CC | (ABP95596-ABP95942) by extracting open-reading frames containing 6-8 |
| CC | transmembrane domains with 250-1000 amino acid residues to give a gene |
| CC | homologous with a known GPCR gene. The receptor proteins and encoded |
| CC | genes are useful for studying in vivo signal transduction mechanism and |
| CC | identifying targets for drug development e.g. based on olfactory and |
| CC | gustatory receptors in form of agonists and antagonists by screening |
| CC | intrinsic and extrinsic ligands as bitter taste inhibitors, taste |
| CC | enhancers and fragrance improvers. |
| CC | Note: The sequence data for this patent did not form part of the printed |
| CC | specification, but was obtained in electronic format directly from WIPO |
| CC | at ftp.wipo.int/pub/published_pct_sequences. |
| XX | |
| SO | Sequence 316 AA; |
| QY | Query Match 1.6%; Score 8; DB 23; Length 316; |
| Db | Best Local Similarity 100.0%; Pred. No. 44; |
| | Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |
| | 94 LFYAVGTP 101 |
| | |
| | 274 LFYAVGTP 281 |
| RESULT 36 | |
| ABJ04015 | |
| ID | ABJ04015 standard; Protein; 316 AA. |
| XX | |
| AC | ABJ04015; |
| XX | |
| DT | 25-SEP-2002 (first entry) |
| DE | |
| XX | Human G-protein coupled receptor SEQ ID NO: 96. |
| XX | |
| KW | Human; G-Protein coupled receptor; GPCR; developmental disease; |
| KW | immune disease; retinal disease; metabolic disorder; feeding disorder; |
| KW | wasting disorder; infection; cancer; allergy; neurological disorder; |
| KW | behavioural disorder; reproductive disorder; antidiabetic; cardiac; |
| XX | antiatherosclerotic. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | WO200255558-A2. |
| XX | |
| PD | 18-JUL-2002. |
| XX | |
| PF | 18-DEC-2001; 2001WO-US49112. |
| XX | |
| XX | |
| PR | 18-DEC-2000; 2000US-256635P. |
| PR | 21-DEC-2000; 2000US-257876P. |
| PR | 04-JAN-2001; 2001US-259743P. |
| PR | 10-JAN-2001; 2001US-260718P. |
| PR | 12-JAN-2001; 2001US-261498P. |
| PR | 24-JAN-2001; 2001US-263689P. |
| PR | 08-FEB-2001; 2001US-267464P. |
| PR | 22-FEB-2001; 2001US-271021P. |
| PR | 14-MAR-2001; 2001US-275946P. |
| PR | 23-MAR-2001; 2001US-278150P. |
| PR | 18-APR-2001; 2001US-284591P. |
| PR | 23-APR-2001; 2001US-285718P. |
| PR | 19-JUN-2001; 2001US-299327P. |
| PR | 16-AUG-2001; 2001US-312902P. |

| | |
|-----------------------|---|
| XX | (CURA-) CURAGEN CORP. |
| PA | |
| XX | |
| PI | Padigaru M, Kekuda R, Li L, Ballinger RA, Caeman SJ, Szytek KA; |
| PI | Colman SD, Vernet CAM, Shenoy S, Guesz V, Malayankar UM, Edinger S, |
| PI | Gerlach V, Smithson G, Stone D, Sciore P, MacDougall JR; |
| PI | Gunther E, Peyman JA, Ellerman K, Millet I, Tchernov VT; |
| PI | Anderson DW, Wolenc A; |
| XX | |
| DR | WPI; 2002-500840/53. |
| DR | N-PSDB; ABT04182. |
| XX | |
| PT | New G-protein coupled receptor polypeptides, encoding nucleic acids and |
| PT | immunospecific antibodies, useful for treating cardiomyopathy, |
| PT | atherosclerosis, diabetes or disorder related to cell signal processing |
| PT | and metabolic pathway modulation - |
| XX | |
| PS | Claim 1; Page 58; 473pp; English. |
| XX | |
| CC | The present invention provides the protein and coding sequences of human |
| CC | G-protein coupled receptors designated GPCR _X . These can be used in the |
| CC | treatment of developmental diseases, immune diseases, retinal diseases, |
| CC | metabolic disorders, feeding disorders, wasting disorders, infections, |
| CC | cancer, allergies, neurological disorders, behavioural disorders and |
| CC | reproductive disorders. The present sequence is a protein of the |
| CC | invention. |
| XX | |
| SQ | Sequence 316 AA; |
| XX | |
| Query March | 1.6%; Score 8; DB 23; Length 316; |
| Best Local Similarity | 100.0%; Pred. NO. 44; |
| Matches 8; | Conservative 0; Mismatches 0; Indels 0; Gaps 0 |
| Oy | 94 LFYAVGTP 101 |
| Db | 274 LFYAVGTP 281 |
| XX | |
| RESULT 37 | |
| ID | ABJ04016 |
| ID | ABJ04016 standard; Protein; 316 AA. |
| AC | ABJ04016; |
| XX | |
| DT | 25-SEP-2002 (first entry) |
| XX | |
| DE | Human G-protein coupled receptor SEQ ID NO: 98. |
| XX | |
| KW | Human; G-protein coupled receptor; GPCR; developmental disease; |
| KW | immune disease; retinal disease; metabolic disorder; feeding disorder; |
| KW | wasting disease; infection; cancer; allergy; neurological disorder; |
| KW | behavioural disorder; reproductive disorder; antidiabetic; cardiac; |
| KW | antiatherosclerotic. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | MO200255558-A2. |
| XX | |
| XX | 18-JUL-2002. |
| PF | 18-DEC-2001; 2001WO-US49112. |
| XX | |
| PR | 18-DEC-2000; 2000US-256635P. |
| PR | 21-DEC-2000; 2000US-257876P. |
| PR | 04-JAN-2001; 2001US-259743P. |
| PR | 10-JAN-2001; 2001US-260718P. |
| PR | 12-JAN-2001; 2001US-261498P. |
| PR | 24-JAN-2001; 2001US-263689P. |
| PR | 08-FEB-2001; 2001US-267464P. |
| PR | 22-FEB-2001; 2001US-271021P. |
| PR | 14-MAR-2001; 2001US-275946P. |
| PR | 23-MAR-2001; 2001US-278150P. |
| PR | 18-APR-2001; 2001US-284591P. |

PR 23-APR-2001; 2001US-285718P.
PR 19-JUN-2001; 2001US-2993227P.
PR 16-AUG-2001; 2001US-312902P.
XX
XX (CURA-) CURAGEN CORP.
XX Padigar M, Kekuda R, Li L, Ballinger RA, Casman SJ, Spytek KA;
PI Colman SD, Vernet CM, Shenoy S, Gusev V, Malpankar UM, Edlinger S;
PI Gerlach V, Smithson G, Stone D, Sciore P, MacDougall JR;
PI Gunther E, Peyman JA, Ellerman K, Millet I, Tchernev VT;
PI Anderson DW, Wolenc A;
XX WPI; 2002-500840/53.
DR N-PSDB; ABT04183.
XX
XX
XX New G-protein coupled receptor polypeptides, encoding nucleic acids and
PT immunospecific antibodies, useful for treating cardiomyopathy,
PT atherosclerosis, diabetes or disorder related to cell signal processing
PT and metabolic pathway modulation -
XX
XX
XX Claim 1; Page 59; 473pp; English.
XX
XX The present invention provides the protein and coding sequences of human
CC G-protein coupled receptors designated GPCR. These can be used in the
CC treatment of developmental diseases, immune diseases, retinal diseases,
CC metabolic disorders, feeding disorders, wasting disorders, infections,
CC cancer, allergies, neurological disorders, behavioural disorders and
CC reproductive disorders. The present sequence is a protein of the
CC invention.
XX
XX Sequence 316 AA;
SQ
Query March 1.6%; Score 8; DB 23; Length 316;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 94 LFYAVGTP 101
Db 274 LFYAVGTP 281
RESULT 38
AAG58322
ID AAG58322 standard; Protein; 317 AA.
XX
XX AAG58322;
AC
XX 18-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 75272.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX EPI033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 25-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142927.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 16-AUG-1999; 99US-0148684.
 PR 17-AUG-1999; 99US-0149368.
 PR 18-AUG-1999; 99US-0149175.
 PR 20-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 30-AUG-1999; 99US-0151080.
 PR 31-AUG-1999; 99US-0151103.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 15-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158359.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161982.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.6%; Score 8; DB 21; Length 317;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 VMLSLMFW 169
 Db 282 VMLSLMFW 289

RESULT 39
 AAG72904
 ID AAG72904 standard; Protein; 318 AA.
 XX
 AC AAG72904;
 XX
 DT 30-JUL-2001 (first entry)
 XX
 DE Human olfactory receptor data exploratorium sequence, SEQ ID NO: 2586.
 XX
 KW Human; olfactory receptor; OR; primary scent determination;
 KW secondary scent determination; polypeptide library; odour receptor;
 KW scent profile; scent fingerprint; scent representation;
 KW human olfactory receptor data exploratorium; HORDE.
 XX
 OS Homo sapiens.
 XX
 PN WO200127158-A2.
 XX
 PD 19-APR-2001.
 XX
 PF 06-OCT-2000; 2000WO-US27562.
 XX
 PR 08-OCT-1999; 99US-0158615.
 PR 24-FEB-2000; 2000US-0184809.
 XX
 PA (DIGI-) DIGISCENTS.
 PA (YEDA) YEDA RES & DEV CO LTD.
 XX
 PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
 XX
 DR WPI; 2001-290713/30.
 XX
 PT New polynucleotides which encode polypeptides involved in olfactory
 PT sensation for identifying olfactory agonists and antagonists -
 XX
 PS Example 6; Page 1746-1747; 1857pp; English.
 XX
 CC The present sequence is a polypeptide from the human olfactory receptor

CC data exploratorium (HORDE). It was used as a query sequence in a database
 CC search of olfactory receptor (OR)-like sequences. The invention relates
 CC to isolated polynucleotides encoding polypeptides involved in olfactory
 CC sensation. The polynucleotides can be used in screening for olfactory
 CC agonists and antagonists. The methods allow for the determination of
 CC primary scents and the identification of the odour receptors used to
 CC detect these primary scents. The methods also enable determination of
 CC secondary scents and the identification of combinations of odour
 CC receptors that are involved in detecting such secondary scents. This
 CC enables the construction of a scent representation (also called a scent
 CC fingerprint or scent profile), which may be used to re-create and edit
 CC scents. Libraries of olfactory receptors are useful for determining the
 CC interaction pattern of a composition with the receptors, and can be
 CC used for determining differences in the olfactory faculties of different
 CC individuals.

CC Sequence 318 AA;

Query Match 1.6%; Score 8; DB 22; Length 318;

Best Local Similarity 100.0%; Pred. No. 44; Mismatches 0; Gaps 0;

Matches 8; Conservative 0; Indels 0; Gaps 0;

Qy 94 LFYAVGTP 101

Db 278 LFYAVGTP 285

RESULT 40

AAAG72622 standard; Protein; 320 AA.

AAAG72622;

31-JUL-2001 (first entry)

Murine OR-like polypeptide query sequence, SEQ ID NO: 2303.

Mouse: olfactory receptor; OR; primary scent determination;

secondary scent determination; polypeptide library; odour receptor;

scent profile; scent fingerprint; scent representation.

Mus musculus.

WO200127158-A2.

19-APR-2001.

06-OCT-2000; 2000WO-US27582.

08-OCT-1999; 99US-0158615.

24-FEB-2000; 2000US-0184809.

(DIGI-) DIGISCENTS.

(YEDA) YEDA RES & DEV CO LTD.

Belenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

WPI; 2001-290713/30.

New polynucleotides which encode polypeptides involved in olfactory

sensation for identifying olfactory agonists and antagonists -

Example 6; Page 1553-1554; 1857pp; English.

The present sequence is a polypeptide encoded by one of 105 newly mined
 CC mouse genes. It was used as a query sequence in a database search of
 CC olfactory receptor (OR)-like sequences. The invention relates to isolated
 CC polynucleotides encoding polypeptides involved in olfactory sensation.
 CC The polynucleotides can be used in screening for olfactory agonists and
 CC antagonists. The methods allow for the determination of primary scents
 CC and the identification of the odour receptors used to detect these
 CC primary scents. The methods also enable determination of secondary scents
 CC and the identification of combinations of odour receptors that are

CC involved in detecting such secondary scents. This enables the
 CC construction of a scent representation (also called a scent fingerprint
 CC or scent profile), which may be used to re-create and edit scents.
 CC Libraries of olfactory receptors are useful for determining the
 CC interaction pattern of a composition with the receptors, and can be
 CC used for determining differences in the olfactory faculties of different
 CC individuals.

CC Sequence 320 AA;

Query Match 1.6%; Score 8; DB 22; Length 320;

Best Local Similarity 100.0%; Pred. No. 44; Mismatches 0; Gaps 0;

Matches 8; Conservative 0; Indels 0; Gaps 0;

Qy 94 LFYAVGTP 101

Db 280 LFYAVGTP 287

Search completed: November 25, 2003, 10:05:20
 Job time : 56 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:08:07 ; Search time 18 Seconds

(without alignments)
1345,486 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 2630
Sequence: 1 MTKTEBKPGFLRSLFWPIH.....AQSLKQGEVAGQEDSAPASS 515

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 2617 | 99.5 | 515 | 1 | Q98J2 chlamydia p |
| 2 | 2147 | 81.6 | 528 | 1 | O84068 chlamydia t |
| 3 | 2143 | 81.5 | 529 | 1 | O9pkx5 chlamydia m |
| 4 | 1333 | 50.7 | 624 | 1 | Q39002 arabidopsis |
| 5 | 1330.5 | 50.6 | 618 | 1 | P92935 arabidopsis |
| 6 | 1306 | 49.7 | 631 | 1 | O24381 solanum tub |
| 7 | 1056 | 40.2 | 498 | 1 | P19568 rickettsia |
| 8 | 1025.5 | 39.0 | 540 | 1 | Q92700 chlamydia p |
| 9 | 969.5 | 36.9 | 501 | 1 | O92670 rickettsia |
| 10 | 956.5 | 36.4 | 543 | 1 | O84502 chlamydia t |
| 11 | 953 | 36.2 | 543 | 1 | O92Jp6 chlamydia m |
| 12 | 880.5 | 33.5 | 512 | 1 | O92d47 rickettsia |
| 13 | 865 | 32.9 | 500 | 1 | O05962 rickettsia |
| 14 | 824 | 31.3 | 507 | 1 | O92df2 rickettsia |
| 15 | 135.5 | 5.2 | 639 | 1 | O48624 leucostoc |
| 16 | 133 | 5.1 | 669 | 1 | P43335 haemophilus |
| 17 | 132.5 | 5.0 | 463 | 1 | P19903 escherichia |
| 18 | 129 | 4.9 | 462 | 1 | NAKU_ECOLI |
| 19 | 128.5 | 4.9 | 528 | 1 | COXI_TRIRU |
| 20 | 126.5 | 4.8 | 512 | 1 | OPUD_BACSU |
| 21 | 126.5 | 4.8 | 526 | 1 | COXI_CYACA |
| 22 | 126 | 4.8 | 664 | 1 | PTSA_STRNU |
| 23 | 125.5 | 4.8 | 641 | 1 | RAFP_PEDPE |
| 24 | 123.5 | 4.7 | 394 | 1 | YMF_F_BACSU |
| 25 | 122 | 4.6 | 687 | 1 | AFUB_ACTPL |
| 26 | 121.5 | 4.6 | 894 | 1 | YNB6_YEAST |
| 27 | 121 | 4.6 | 536 | 1 | COXI_ALAMA |
| 28 | 120 | 4.6 | 389 | 1 | YDHP_ECOLI |
| 29 | 119 | 4.5 | 389 | 1 | YDHP_ECO57 |
| 30 | 118 | 4.5 | 462 | 1 | NAKU_SALTU |
| 31 | 117.5 | 4.5 | 504 | 1 | POTP_HAEIN |
| 32 | 116.5 | 4.4 | 522 | 1 | COXI_MARPO |
| 33 | 116 | 4.4 | 635 | 1 | NRFE_HAEIN |

| | | | | | | |
|-----|-------|-----|------|---|------------|---------------------|
| 34 | 115.5 | 4.4 | 397 | 1 | YGED_ECOLI | P39196 escherichia |
| 35 | 115 | 4.4 | 521 | 1 | COXI_APIRI | P20374 apis mellif |
| 36 | 115 | 4.4 | 541 | 1 | COXI_PODAN | P20681 podospora a |
| 37 | 115 | 4.4 | 562 | 1 | COXI_THERM | O56408 thermus the |
| 38 | 115 | 4.4 | 873 | 1 | COXI_ACACA | O37370 acanthamoeb |
| 39 | 114.5 | 4.4 | 380 | 1 | CYB_PINIM | O81w83 piniguns i |
| 40 | 114.5 | 4.4 | 593 | 1 | Y917_MYCTU | O05909 mycobacteri |
| 41 | 114 | 4.3 | 557 | 1 | ATKA_ECOL6 | O86fj3 escherichia |
| 42 | 113.5 | 4.3 | 461 | 1 | Y608_HAEIN | O57486 haemophilus |
| 43 | 112.5 | 4.3 | 379 | 1 | CYB_BALBN | P41281 balaenopter |
| 44 | 112.5 | 4.3 | 393 | 1 | SRPC_SYNP7 | O50277 synechococc |
| 45 | 111.5 | 4.2 | 380 | 1 | CYB_ALALA | O86x23 alle alle (|
| 46 | 111 | 4.2 | 557 | 1 | ATKA_ECO57 | O8x9f8 escherichia |
| 47 | 111 | 4.2 | 502 | 1 | COXI_EMENT | P00402 emericella |
| 48 | 110.5 | 4.2 | 405 | 1 | YUHB_ECOLI | P39352 escherichia |
| 49 | 110.5 | 4.2 | 530 | 1 | COXI_MERSE | O35101 metridium s |
| 50 | 110.5 | 4.2 | 557 | 1 | ATKA_ECOLI | P39959 escherichia |
| 51 | 110 | 4.2 | 447 | 1 | Y418_HAEIN | P44699 haemophilus |
| 52 | 109.5 | 4.2 | 379 | 1 | CYB_LAMVI | O36227 lama vicugn |
| 53 | 109.5 | 4.2 | 502 | 1 | NAT2_ECOL6 | O86fdu escherichia |
| 54 | 109.5 | 4.2 | 1107 | 1 | YUHP_ECOLI | P39285 escherichia |
| 55 | 109 | 4.1 | 380 | 1 | CYB_HARDI | O79520 harpactes d |
| 56 | 108.5 | 4.1 | 379 | 1 | CYB_BALPH | P24950 balaenopter |
| 57 | 108.5 | 4.1 | 460 | 1 | NU4M_BRAPR | O9miy1 brachydanio |
| 58 | 108.5 | 4.1 | 532 | 1 | COXI_CHOCR | P48866 chondrus cr |
| 59 | 108 | 4.1 | 387 | 1 | CYB_FODAN | O02655 podospora a |
| 60 | 108 | 4.1 | 433 | 1 | DCUA_ECOLI | P44539 escherichia |
| 61 | 107.5 | 4.1 | 439 | 1 | CYB_MEGNO | P41289 megaptera n |
| 62 | 107.5 | 4.1 | 548 | 1 | SHIA_ECOLI | P76350 escherichia |
| 63 | 107 | 4.1 | 438 | 1 | YJCG_ECOLI | P32705 escherichia |
| 64 | 107 | 4.1 | 555 | 1 | CSTA_AQUAE | O67304 aquifex aeo |
| 65 | 107 | 4.1 | 658 | 1 | S282_HUMAN | O43868 homo sapien |
| 66 | 106.5 | 4.0 | 380 | 1 | CYB_ALCTO | O8m0a6 alca lorda |
| 67 | 106.5 | 4.0 | 380 | 1 | CYB_URILLO | O81wma uria lomvia |
| 68 | 106.5 | 4.0 | 492 | 1 | YD19_METUA | O58715 methanococc |
| 69 | 106.5 | 4.0 | 509 | 1 | YFCC_HAEIN | P44023 haemophilus |
| 70 | 106.5 | 4.0 | 514 | 1 | COXI_PROMI | O05143 prototheca |
| 71 | 106.5 | 4.0 | 524 | 1 | COXI_BERVU | P24794 beta vulgar |
| 72 | 106.5 | 4.0 | 527 | 1 | NU4M_ANOGA | COXI_ARATH |
| 73 | 106 | 4.0 | 447 | 1 | COXI_ARATH | O07063 arabidopsis |
| 74 | 105.5 | 4.0 | 380 | 1 | CYB_GYMTI | P34852 anopheles g |
| 75 | 105.5 | 4.0 | 380 | 1 | CYB_URIAL | P29637 gymnomorpha |
| 76 | 105.5 | 4.0 | 383 | 1 | Y942_SYNV3 | O81wmd uria aalge |
| 77 | 105.5 | 4.0 | 388 | 1 | SH4_HUMAN | P74311 synechocyst |
| 78 | 105.5 | 4.0 | 394 | 1 | ARAU_ECOLI | Q13638 homo sapien |
| 79 | 105.5 | 4.0 | 524 | 1 | COXI_ORYSA | P33910 escherichia |
| 80 | 105.5 | 4.0 | 524 | 1 | COXI_WHEAT | P14578 oryza sativ |
| 81 | 105.5 | 4.0 | 528 | 1 | COXI_MAIZE | P08741 triticum ae |
| 82 | 105.5 | 4.0 | 530 | 1 | COXI_SORBI | P08742 zea mays (m |
| 83 | 105.5 | 4.0 | 622 | 1 | COXI_BACSU | P05502 sorghum bic |
| 84 | 105 | 4.0 | 336 | 1 | AGLF_RHIME | P24010 bacillus su |
| 85 | 105 | 4.0 | 416 | 1 | LACY_KLEPM | Q92356 rhizobium m |
| 86 | 105 | 4.0 | 468 | 1 | CLCA_VIBCH | P18817 klebsiella |
| 87 | 105 | 4.0 | 500 | 1 | YJEM_ECOLI | Q9kme2 vibrio chol |
| 88 | 105 | 4.0 | 548 | 1 | TOXA_COCCA | P92862 escherichia |
| 89 | 105 | 4.0 | 569 | 1 | SOTI_SPIOL | O00357 cochlilobolu |
| 90 | 104.5 | 4.0 | 342 | 1 | Y762_METUA | Q11364 spinacia ol |
| 91 | 104.5 | 4.0 | 379 | 1 | CYB_IAMGL | O58172 methanococc |
| 92 | 104.5 | 4.0 | 380 | 1 | CYB_BUTBU | Q44891 lama glama |
| 93 | 104.5 | 4.0 | 380 | 1 | CYB_NUMME | Q35160 buteo buteo |
| 94 | 104.5 | 4.0 | 461 | 1 | PUCG_RHOCA | O35160 numida mele |
| 95 | 104.5 | 4.0 | 534 | 1 | COXI_RICPR | P33462 rhodobacter |
| 96 | 104.5 | 4.0 | 741 | 1 | YBIO_ECOLI | O54066 rickettsia |
| 97 | 104 | 4.0 | 500 | 1 | NU4C_NEPOL | P75783 escherichia |
| 98 | 104 | 4.0 | 587 | 1 | S132_RAT | Q6kxv8 nephrolelmi |
| 99 | 104 | 4.0 | 1120 | 1 | KEPA_ECOLI | P70545 ratus norv |
| 100 | 103.5 | 3.9 | 368 | 1 | Y812_AQUAE | P73318 escherichia |
| | | | | | | O66994 aquifex aeo |

ALIGNMENTS

RESULT 1

TLCL1_CHLTPN STANDARD; PRT; 515 AA.
 AC Q928J2; Q9USF8; 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP/ATP carrier protein 1 (ADP/ATP translocase 1).
 GN TLCA OR ADP1 OR CPN0351 OR CP0408.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.
 NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CHL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Baas S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shira T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RA Shira T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CHL029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE001619; AAD18495.1; -
 CC EMBL; AE002202; AAF38252.1; -
 CC EMBL; AP002546; BAA98559.1; -
 CC PIR; E72089; E72089.
 CC PIR; E86534; E86534.
 CC TIGR; CP0408; -
 CC InterPro; IPR004667; ADP_ATP_car.
 CC Pfam; PF03219; TLC; 1.
 CC TIGRfam; TIGR00769; AAA; 1.
 CC Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 FT TRANSMEM 24 44 POTENTIAL.
 FT TRANSMEM 62 82 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 124 144 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 226 246 POTENTIAL.
 FT TRANSMEM 286 306 POTENTIAL.
 FT TRANSMEM 329 349 POTENTIAL.
 FT TRANSMEM 358 378 POTENTIAL.
 FT TRANSMEM 383 403 POTENTIAL.

FT TRANSMEM 465 485 POTENTIAL.
 FT CONFLICT 141 141 V -> D (IN REF. 3).
 SQ SEQUENCE 515 AA; 57197 MM; E3F4A52CE008E106 CRC64;
 Query Match 99.5%; Score 2617; DB 1; Length 515;
 Best Local Similarity 99.6%; Pred. No. 5.6e-158;
 Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MTKEEKPFGKLRSEFLPHTHEKLVLPFMEFCTTNTVTYVDRDRTIIVAPSGA 60
 DB 1 MTKEEKPFGKLRSEFLPHTHEKLVLPFMEFCTTNTVTYVDRDRTIIVAPSGA 60
 QY 61 EAIPFIKFWLVPCATIFMLIVAKLSNLSKQALFYVGPFLIFPFIYPLDVL 120
 DB 61 EAIPFIKFWLVPCATIFMLIVAKLSNLSKQALFYVGPFLIFPFIYPLDVL 120
 QY 121 HPTFADRLQAILPPGLGLVAILRNWTPAFVYLAELWGSVMSLMFPGFANEITKHE 180
 DB 121 HPTFADRLQAILPPGLGLVAILRNWTPAFVYLAELWGSVMSLMFPGFANEITKHE 180
 QY 181 AKRFYALFGIGANISLASGRATYVASKTRASVSEGVDPGICISRLMAMTIVSGLVMA 240
 DB 181 AKRFYALFGIGANISLASGRATYVASKTRASVSEGVDPGICISRLMAMTIVSGLVMA 240
 QY 241 SYWINKNVLTPDFYVPEEMQKKGAKPRMKNKDSFLYIDRSPYLLTLIVAYGIC 300
 DB 241 SYWINKNVLTPDFYVPEEMQKKGAKPRMKNKDSFLYIDRSPYLLTLIVAYGIC 300
 QY 301 INLIETWKSQKLQYNNNDYSEFGNFSFWTGVSVLIMLVGVGNVIRKFGMLTGALV 360
 DB 301 INLIETWKSQKLQYNNNDYSEFGNFSFWTGVSVLIMLVGVGNVIRKFGMLTGALV 360
 QY 361 TPVWVLLTGIVFPALVIFRQASGLVAMFGTPTMLAVMGATONITSKTKVLPSTK 420
 DB 361 TPVWVLLTGIVFPALVIFRQASGLVAMFGTPTMLAVMGATONITSKTKVLPSTK 420
 QY 421 EMAYIPDQOKYKGAIVVAFKSGSGALIQGLVYICSGISGAMTPLYAILLFI 480
 DB 421 EMAYIPDQOKYKGAIVVAFKSGSGALIQGLVYICSGISGAMTPLYAILLFI 480
 QY 481 AIWVSAATKLNKFLAQSALKEQVAGDSAPASS 515
 DB 481 AIWVSAATKLNKFLAQSALKEQVAGDSAPASS 515
 RESULT 2
 TLCL1_CHLTPN STANDARD; PRT; 528 AA.
 AC Q928J2; Q9USF8; 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE ADP/ATP carrier protein 1 (ADP/ATP translocase 1).
 GN TLCA OR CT065.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.
 NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UW-3/Ck;
 RX MEDLINE=9900809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AE001281; AAC67656.1; -
 CC PIR; C71561; C71561.
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TIC; 1.
 DR TIGRfams; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KM Complete proteome.
 FT TRANSMEM 24 44 POTENTIAL.
 FT TRANSMEM 63 83 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 124 144 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 220 240 POTENTIAL.
 FT TRANSMEM 284 304 POTENTIAL.
 FT TRANSMEM 327 347 POTENTIAL.
 FT TRANSMEM 356 376 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 463 483 POTENTIAL.
 SO SEQUENCE 528 AA; 58117 MW; EAD16A43B95B94B2 CRC64;

Query Match 81.6%; Score 2147; DB 1; Length 528;
 Best Local Similarity 78.1%; Pred. No. 2.3e-128;
 Matches 410; Conservative 52; Mismatches 49; Indels 14; Gaps 3;

QY 1 MTKTEKRPFGKLSFLPFIHTEHKKVLPMLFFCITPNTYVLRDTKDTLIIVAPSGA 60
 DB 1 MTQTAERPFGLRSLFPIHHEHKKVLPMLFFCISFNITLIRDTKDTLIIVAPSGA 60
 QY 61 EAIPIKFWLVPCALIFMLIYAKLSNLSKQALFYAVGTEFLFAFPVIVPLRDVL 120
 DB 61 EAIPIKFWLVPCALIFMLIYAKLSNLSKQALFYAVGTEFLFAFPVIVPLRDVL 120
 QY 121 HPTFAFADTLQALIPSGGLGVAIIRNMTFAFYVLAELGWSVMSLMFWGFANEITKHE 180
 DB 121 HPTFAFADTLQALIPSGGLGVAIIRNMTFAFYVLAELGWSVMSLMFWGFANEITKHE 180
 QY 121 HPTFAFADTLQALIPSGGLGVAIIRNMTFAFYVLAELGWSVMSLMFWGFANEITKHE 180
 DB 121 HPTFAFADTLQALIPSGGLGVAIIRNMTFAFYVLAELGWSVMSLMFWGFANEITKHE 180
 QY 181 AKRFYALFGIGANTISLASGRAIYVASKLRASVSGVDPWGISRLMAAMTIVGLVMA 240
 DB 181 AKRFYALFGIGANTISLASGRAIYVASKLRASVSGVDPWGISRLMAAMTIVGLVMA 240
 QY 241 SYMMINKNVLTDPFYPNEEMQKGGKAPKMNKDSFLYIDRSPYILLTLVAVGIC 300
 DB 241 SYMMINKNVLTDPFYPNEEMQKGGKAPKMNKDSFLYIDRSPYILLTLVAVGIC 300
 QY 241 CYMMNKRVLTDPRFYPNAPL-KAKK-SKPKMSGSESYLLRSPYMLLALLVICYIC 298
 DB 241 CYMMNKRVLTDPRFYPNAPL-KAKK-SKPKMSGSESYLLRSPYMLLALLVICYIC 298
 QY 301 INLEVTWKSQKLQYPMNDYSEFMGNFSFMTGVSVYLIMLVGVGNVIRKFGMLTGALV 360
 DB 299 INLEVTWKSQKLQYPMNDYSEFMGNFSFMTGVSVYLIMLVGVGNVIRKFGMLTGALV 358
 QY 361 TPWVAVLTGLYFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSKRYALPSTK 420
 DB 359 TPWVAVLTGLYFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSKRYALPSTK 418
 QY 421 EMAYIPLDQOKVKGKALIDVVAARFGSGGALIOGGLVIGSIGAMTPIAVILPIT 480
 DB 419 EMAYIPLDQOKVKGKALIDVVAARFGSGGALIOGGLVIGSIGAMTPIAVILPIT 478
 QY 481 AIWIVSATKINKLPLAOSALKEQEAQ-----EDSAPA 513
 DB 479 MWVLTSAKINKLFLAASAKQELFAAFAAAKEKASAKESAPA 523

RESULT 3
 TLCL CHLMT STANDARD; PRT; 529 AA.
 ID TLCL CHLMT
 AC 09PKS5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP_ATP carrier protein 1 (ADP/ATP translocase 1).
 GN TLCA OR TC0335.
 OS Chlamydia muridarum.
 CC Bacteria; Chlamydiales; Chlamydiales; Chlamydiaceae; Chlamydia.
 CC NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M09N / Ni99;
 RX MEDLINE=20150235; PubMed=10684935;
 RA Read T.D., Brumham R.C., Shen C., Gull S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utecherback T., Berry K., Baas S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Eichen M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.,
 RT "Genome sequences of Chlamydia trachomatis M09N and Chlamydia
 RL pneumoniae AR39.";
 CC Nucleic Acids Res. 28:1397-1406(2000).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TIC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AE002301; AAF39198.1; -
 CC PIR; C81714; C81714.
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TIC; 1.
 DR TIGRfams; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KM Complete proteome.
 FT TRANSMEM 24 44 POTENTIAL.
 FT TRANSMEM 63 83 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 124 144 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 220 240 POTENTIAL.
 FT TRANSMEM 284 304 POTENTIAL.
 FT TRANSMEM 322 342 POTENTIAL.
 FT TRANSMEM 356 376 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 463 483 POTENTIAL.
 SO SEQUENCE 529 AA; 58304 MW; EED13A4C751071C9 CRC64;

Query Match 81.5%; Score 2143; DB 1; Length 529;
 Best Local Similarity 78.5%; Pred. No. 4.1e-128;
 Matches 408; Conservative 55; Mismatches 49; Indels 8; Gaps 3;

QY 1 MTKTEKRPFGKLSFLPFIHTEHKKVLPMLFFCITPNTYVLRDTKDTLIIVAPSGA 60
 DB 1 MTQTAERPFGLRSLFPIHHEHKKVLPMLFFCISFNITLIRDTKDTLIIVAPSGA 60
 QY 61 EAIPIKFWLVPCALIFMLIYAKLSNLSKQALFYAVGTEFLFAFPVIVPLRDVL 120
 DB 61 EAIPIKFWLVPCALIFMLIYAKLSNLSKQALFYAVGTEFLFAFPVIVPLRDVL 120
 QY 121 HPTFAFADTLQALIPSGGLGVAIIRNMTFAFYVLAELGWSVMSLMFWGFANEITKHE 180
 DB 121 HPTFAFADTLQALIPSGGLGVAIIRNMTFAFYVLAELGWSVMSLMFWGFANEITKHE 180
 QY 121 HPTFAFADTLQALIPSGGLGVAIIRNMTFAFYVLAELGWSVMSLMFWGFANEITKHE 180
 DB 121 HPTFAFADTLQALIPSGGLGVAIIRNMTFAFYVLAELGWSVMSLMFWGFANEITKHE 180
 QY 181 AKRFYALFGIGANTISLASGRAIYVASKLRASVSGVDPWGISRLMAAMTIVGLVMA 240
 DB 181 AKRFYALFGIGANTISLASGRAIYVASKLRASVSGVDPWGISRLMAAMTIVGLVMA 240
 QY 241 SYMMINKNVLTDPFYPNEEMQKGGKAPKMNKDSFLYIDRSPYILLTLVAVGIC 300
 DB 241 SYMMINKNVLTDPFYPNEEMQKGGKAPKMNKDSFLYIDRSPYILLTLVAVGIC 300

DB 241 CYMMNRVYLDPRFRYPNPAEL-KAKK-SKPKMKGESPSYLLRSPMLLALVLCYVC 298

QY 301 INLEIETWKSQKLOQYPMNNDYSEFMGNFSEWTGVVSLIMLFVCGNIRKFGMTGLV 360

DB 299 INLEIETWKSQKLOQYPMNNDYSEFMGNFSEWTGVVSLIMLFVCGNIRKFGMTGLV 358

QY 361 TPVAVLTLGIYEFALVIFRNQSGVAWFGTPLMLAVVGAIONISKSTFYALFSDTK 420

DB 359 TPVAVLTLGIYEFALVIFRNQSGVAWFGTPLMLAVVGAIONISKSTFYALFSDTK 418

QY 421 EMAYIPLDQEKVKGAIDVVAARFGKSGALIOGGLVIGSIGAMTPYLAIVLLEPI 480

DB 419 EMAYIPLDQEKVKGAIDVVAARFGKSGALIOGGLVIGSIGAMTPYLAIVLLEPI 478

QY 481 AIVLVSATKLNKLFIAQSALKRQEV-----AQESAPAS 514

DB 479 MVWLTSATKLNKLFIAQSALKRQEV-----AQESAPAS 518

RESULT 4

TLCL ARATH STANDARD; PRT; 624 AA.

ID TLCL ARATH STANDARD; PRT; 624 AA.

AC Q39002; Q94AN9; Q9C974; (Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last sequence update)

DE Chloroplast ADP/ATP carrier protein 1, chloroplast precursor (ADP/ATP translocase 1) (Adenine nucleotide translocase 1).

GN AATP1 OR AT1G80300 OR F516.5.

OS Arabidopsis thaliana (Mouse-ear cress).

CC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucotsids II; Brassicales; Brassicaceae; Arabidopsids.

OC NCBI TaxID=3702;

OX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia; TISSUE=Seedling;

EX MEDLINE=96069943; PubMed=7589569;

RA Kampfenkel K., Moehlmann T., Batz O., van Montagu M., Inze D., Neuhaus H.E.;

RT "Molecular characterization of an Arabidopsis thaliana cDNA encoding a novel putative adenylate translocator of higher plants.";

RL FEBS Lett. 374:351-355(1995).

RN [2]

RP REVISIONS.

RA Kampfenkel K.;

RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

EX MEDLINE=21016719; PubMed=1130712;

RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y., Chughler E., Chen Q., Chen R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K., Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y., Gill J.E., Goldsmith A.D., Haase B., Hansen N.F., Hughes B., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A., Miltchev J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I., Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., San H., Salzborg S.L., Schwartz J.R., Shim P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D., Wu D., Yu G., Fraser C.M., Vaysberg M., Vysotskaia V.S., Walker M., "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";

RT Nature 408:816-820(2000).

RL (4)

RN SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;

RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the SSP consortium (Salk/Stanford/PGEC).";

RT Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.

RL [5]

RN FUNCTION, AND SUBCELLULAR LOCATION.

RP MEDLINE=97117797; PubMed=9025303;

RA Neuhaus H.E., Thom E., Mohlmann T., Steup M., Kampfenkel K.;

RT "Characterization of a novel eukaryotic ATP/ADP translocator located in the plastid envelope of Arabidopsis thaliana L.";

RL Plant J. 11:73-82(1997).

CC -1- FUNCTION: May function as an ATP importer.

CC -1- SUBCELLULAR LOCATION: CHLOROPLAST. INTEGRAL MEMBRANE PROTEIN.

CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; Z49227; CAAB9201.2; -.

DR EMBL; AC018848; AAG52434.1; -.

DR EMBL; AF428316; AA16246.1; -.

DR EMBL; AY045903; AAK76577.1; -.

DR PIR; E36834; E36834; ADP_ATP_car.

DR InterPro; IPR004667; ADP_ATP_car.

DR Pfam; PF03219; TLC; 1.

DR TIGRFAMs; TIGR00769; AAA; 1.

KW Transmembrane; Transport; ATP-binding; Multigene family;

KW Chloroplast; Transit peptide.

FT TRANSIT 1 624

FT CHAIN ? 624

FT TRANSMEM 108 128

FT TRANSMEM 182 202

FT TRANSMEM 240 260

FT TRANSMEM 315 335

FT TRANSMEM 446 466

FT TRANSMEM 545 565

FT CONFLICT 45 94

FT CONFLICT 103 103

FT CONFLICT 197 197

FT CONFLICT 281 281

FT CONFLICT 624 AA; 530302A89AD08745 CRC64;

SEQ

Query Match 50.7%; Score 1333; DB 1; Length 624;

Best local similarity 51.8%; Pred. No. 4.9e-77;

Matches 259; Conservative 89; Mismatches 140; Indels 12; Gaps 3;

QY 16 LMPHITHELKVLDPMLFMFCITFNVTYVLRDQTLIVGAPGSAEAIPIKFMVLVPCA 75

DB 105 IFGEVAVTLTKIIPGLGMFCILFNVTYVLRDQTLIVGAPGSAEAIPIKFMVLVPCA 164

QY 76 IIPMLITAKSLNISKQALPFAVGTPLPILFALPPYIYPRADVLHPRFDRQALLPP 135

DB 165 IGFMLLYTKSNVLSKSKALFTVTVIPPIYFGAGFWVPLSNYIHPEALDKLITLTPG 224

QY 136 GLIGLVILRWTPAAAFVLAELGWSVLSLMPFGPANEITKIHAKRFYALFGIGANIS 195

DB 225 RFKQPIALIRMSCLTYVAELMGSVVSVLFPFGANQITTVDEAKKFYLPFGIGANVA 284

QY 196 LLAAGRAIWAASKLRASVSEGVDPWGISLRLLAMTVTSGVILMASTWINKNVLTDPF 255

DB 285 LIFGRITVYKYSNRKMLPGVDGMVSLKMMMSIVGGMGLICLLYWMVVRVYPLPTRS 344

QY 256 YNPEEMOKKGAAPKPMKMSFLYLDSPYIILLTLVLAAYGICVILIEVTKSQKLO 315

DB 345 KKKKE-----KPKMGWESLKFVSSPPIRLDLATLVAYGISTINLEVTWKSQKLO 396

QY 316 YPNNDYSEFMGNFSEWTGVVSLIMLFVCGNIRKFGMTGLVLTVPVAVLTLGIYEFAL 375

```

Db      397 FSPSENEYSAGFMGDFSTCGATFTMTL-LSQYFENKYGWGAAKITPTVLLTGVAFPSL 455
Qy      376 VIFPNQASGLVAMGCTPMLAVVGAIONISKSTKALPDSTKEMAYIPLDQOKYK 435
Db      456 ILFGGPPALVAKIGMTPLAAVVGALONIFSSAKYSLDPCKEMAYIPLDDETKYK 515
Qy      436 KAAIDVVAARFGKSGGALIOGGLVITGSGIGAMPYLAIVLLFITAIVLSATKLNKLEL 495
Db      516 KAAIDVVCNPLGKSGGALIOGPMILSFGSLANSTPYLGMILLVITAMLAAKSLGEGF- 574
Qy      496 AQSALKEQEVAAQEDSAPASS 515
Db      575 --NSLRSEELEREMERASS 592

RESULT 5
TLC2_ARATH STANDARD; PRT; 618 AA.
ID      TLC2_ARATH
AC      P92935; GQM9E3;
DT      30-MAY-2000 (Rel. 39, Created)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Chloroplast ADP/ATP carrier protein 2, chloroplast precursor (ADP/ATP
translocase 2) (Adenine nucleotide translocase 2).
GN      ATP2 OR AT1G15500 OR T16N11.1.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98206726; PubMed=9546649;
RA      Moehlmann T., Thaden U., Schwoeppe C., Winkler H.H., Kampfenkel K.,
RA      Neuhaus H.E.;
RT      "Occurrence of two plastidic ATP/ADP transporters in Arabidopsis
thaliana L. -- molecular characterisation and comparative structural
analysis of similar ATP/ADP translocators from plastids and
RT      Rickettsia prowazekii."
RL      Eur. J. Biochem. 252:353-359 (1998).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      STRAIN=cv. Columbia;
RA      MEDLINE=21016719; PubMed=11130712;
RA      Theologis A., Ecker J.R., Palm C.J., Federepiel N.A., Kaul S.,
RA      White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
RA      Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA      Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA      Dunn P., Egu P., Feldlyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA      Gali J.E., Goldenrich A.D., Haas B., Hansen N.P., Hughes B., Huzar L.,
RA      Hunter J.L., Jenkins U., Johnson-Hopson C., Khan S., Khaykin E.,
RA      Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA      Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA      Lin X., Liu S.X., Liu Z.A., Lueros J.S., Malit R., Marzilli A.,
RA      Miltchev J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,
RA      Pei G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA      Sakano H., Salzborg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA      Sun H., Tallon L.B., Tambunga G., Toriumi M.J., Town C.D.,
RA      Utechtack T., Van Aken S., Vayenberg M., Vayetskaia V.S., Walker M.,
RA      Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT      "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana."
RL      Nature 406:816-820 (2000).
RN      [3]
RP      SEQUENCE FROM N.A.
RX      STRAIN=cv. Columbia;
RA      Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;
RT      "Riken Arabidopsis full length cDNA clones (RATLs) sequenced by the
RT      SSP consortium (Salk/Stanford/PGSC)."
RL      Submitted (JUL-2002) to the EMBL/Genbank/DBJ databases.
RN      [4]
RP      SEQUENCE FROM N.A.

```

```

RA      Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA      Feldmann K.;
RL      Full-length cDNA from Arabidopsis thaliana."
RT      Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
CC      -1- SUBCELLULAR LOCATION: CHLOROPLAST. INTEGRAL MEMBRANE PROTEIN.
CC      -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
CC      -1- CAUTION: Ref.1 sequence differs from that shown due to
CC      framehifts and other sequence problems.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X94626; CAA64329.1; ALT SEQ.
DR      EMBL; AC013453; AAF71976.1; -
DR      EMBL; AY081350; AAL91239.1; -
DR      EMBL; AY128844; AAM91244.1; -
DR      EMBL; AY084374; AAM60955.1; -
DR      PIR; G86288; G86288.
DR      InterPro; IPR004667; ADP_ATP_car.
DR      Pfam; PF03219; TLC; 1.
DR      TrnRFams; TIGR00769; AAA; 1.
DR      TrnSFams; TIGR00769; AAA; 1.
KW      Transmembrane; Transport; ATP-binding; Multigene family;
KW      Chloroplast; Transl peptid.
FT      TRANSLAT 1
FT      CHAIN 1
SQ      SEQUENCE 618 AA; 67530 MM; B2D49E77F17A887A CRC64;

```

```

Query Match 50.6%; Score 1330.5; DB 1; Length 618;
Beet Local Similarity 50.9%; Pred. No. 6.9e-77;
Matches 260; Conservative 92; Mismatches 136; Indels 23; Gaps 5;

Qy      16 LMPHTHELKLVLPFLMFCITPTNTVLRDTKTLVIGAGSGAEMAPFKFWLVVCA 75
Db      102 IFGVAVTLTKIVLGLMFCILENTITLDTKVLVVTAKSSAEIIPFKTWVNLPM 161
Qy      76 IIFMLIYAKLSNLSKQALFYAVGTPLIFPALFPVTVPLRDVLPHTPEADRLQALIP 135
Db      162 IGFMLYTKLSNVLSKQALFYVTVVPLVFGAGFWVPLSNLIHDEALADKLATLGP 221
Qy      136 GLGLGVAILRNWTAAYVLAELMGVMSLMFGPANEITKIHAKRFYALFGIGANIS 195
Db      222 RFMGPLAIMRWSPCLFYVVAELMGVSVVSLFWGFANQITVDEAKKFFPLFGIGANVA 281
Qy      196 LLASGRAIVASKRLRAVSGVDPMGSLRLMAMTIVSGIVMAVYMWINKVLTDPFR 255
Db      282 LIFSGRTVKTFPSNRKNLGPVDSMAVSLKAMSVIGMGALICFLYMWVNR----- 333
Qy      256 YNPEEMOKGKKGAKPKAMNMDSPLYLDRSPYIILTLVIAVGCINLIETVWSQKIQ 315
Db      334 YVLPPTSKKKKKVAPQGMWESLKFIVSSPYIRDLATLVAVGSIINLVETWWSKKAQ 393
Qy      316 YPNMNDYSEFMNGSFPTGVSVLIMLVGGANVIRKGMLTGALVTPVMVLLTGIYFPAL 375
Db      394 FSPSENEYSAGFMGDFSTCGATFTMTL-LSQYFENKYGWGAAKITPTVLLTGVAFPSL 452
Qy      376 VIFPNQASGLVAMGCTPMLAVVGAIONISKSTKALPDSTKEMAYIPLDQOKYK 435
Db      453 ILFGGPPALVAKIGMTPLAAVVGALONIFSSAKYSLDPCKEMAYIPLDDETKYK 512
Qy      436 KAAIDVVAARFGKSGGALIOGGLVITGSGIGAMPYLAIVLLFITAIVLSATKLNKLEL 491
Db      513 KAAIDVVCNPLGKSGGALIOGPMILSFGSLANSTPYLGMILLVITAMLAAKSLGEGFN 572
Qy      492 KFLAQSALKQF-----VAQEDSAPA 513
Db      573 TLMESEELEREMERASSVKIPVVSQED-APS 602

```

RESULT 6
 TLCL SOLTU STANDARD; PRT; 631 AA.
 AC O24381;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Plastidic ATP/ADP-transporter.
 OS Solanum tuberosum (Potato).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Asceridae; lamids; Solanales; Solanaceae; Solanum.
 CC NCBI_TaxID=4113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Desiree;
 RA Neuhaus E.;
 RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST. INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Y10821; CAA1785.1; -
 DR PIR; T07420; T07420.
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC; 1.
 DR TrEMBL; T1GR00769; AAA; 1.
 DR Transmembrane; Transport; ATP-binding; Multigene family;
 KW Chloroplast.
 SQ SEQUENCE 631 AA; 68886 MW; 33158380DE2901EA CRC64;
 Query Match 49.7%; Score 1306; DB 1; Length 631;
 Best Local Similarity 52.1%; Pred. No. 2.5e-75;
 Matches 256; Conservative 87; Mismatches 136; Indels 12; Gaps 3;
 Oy 24 LKKVLPMLPFCEITFENYTVLRDTDTLIVGAPSGAGAEIPRIKFWLVPCCIIIFMLIYA 83
 Db 111 LKTIIPDLMFPCIIIFNTITLIDTDLVLTAKSSAEIIPFKIWMVDPMAIGMLIYT 170
 Oy 84 KLSNLSKQALFYAVGTPELIFPFIYVYPLRVLPHTPEFADRLQALPPGLGLVAI 143
 Db 171 KLANVLSKQALFYVITLFPFIYVYPLRVLPHTPEFADRLQALPPGLGLVAI 230
 Oy 144 LRNMTFAAFYVLAELMGVSLMGMWGRANETTKHEAKRFYALFGIGANISLASSRAI 203
 Db 221 LRMSFCIFYVAELMGVSVLVLPWGFANQITTVDEKRFYPLGLGANVALIFSGRTV 290
 Oy 204 VMASGLRASVSGVDPWGISLRLLAMNTISGLVVMASVYMNKNVLDPRRYNDEEMQK 263
 Db 291 KYFSSLRSLSGPGVGMALISLKGMMSIYVMGGAICFYRWVNRVVALPTR-----S 342
 Oy 264 GKKGAKPRNNKDSFLYLRSPYILLTLTLVAVGICINLIEVTKSOLKLYPMNDYS 323
 Db 343 KKKKVKPRNMTMESLKFVSSKYINDLATLVAVGISINLVETWKSGLKQAFPNESYS 402
 Oy 324 EPMGNFSGTWGVSVLMLFVGNVIRKFGMLTGALVTPVWVLTLGIYFPAIVIRNQS 383
 Db 403 SFMGDFSTATGATPTMML-LSQWFFDYXGWAAKIPITVILLTGVGFFSLIFGABLA 461
 Oy 384 GLVMEFGTPTMLAVVAVGAIQNLISKTKVALFDSKEMAYPLDOEKVKGKAIDVYA 443
 Db 462 PTLAFTGPTPLAAAYVGMQNTFSKAKYSIFDCKEMAYIPLDEDVTKGKAIDVVC 521
 Oy 444 ARFGSGGALLQOGLLVICSGIGANTPYAVLVLLFIILVIVLSATKNKLPFAGSALKEQ 503
 Db 522 NPLKSGGALLQOFPVILTFGSLASTPYLGGVLIVIVLWAGAAKSLDGGF---TQLNGE 578

Oy 504 EVAQEDSAPAS 514
 Db 579 EDLEKEMERAS 589
 RESULT 7
 TLCA RICPR STANDARD; PRT; 498 AA.
 AC P19568;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP ATP carrier protein 1 (ADP/ATP translocase 1).
 CC TLCA OR TLCL OR TLCP OR RP053.
 OS Rickettsia prowazekii.
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 CC NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=255259;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria."
 RL Nature 396:133-140 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=91258319; PubMed=1904433;
 RA Plano G.V., Winkler H.H.;
 RT "Identification and initial topological analysis of the Rickettsia
 RT prowazekii ATP/ADP translocase."
 RL J. Bacteriol. 173:3389-3396 (1991).
 RN [4]
 RP TOPOLOGY.
 RX MEDLINE=90334349; PubMed=2165754;
 RA Plano G.V., Wood D.O., Winkler H.H.;
 RT "Rickettsia prowazekii and ATP/ADP translocase. Analysis of gene
 RT fusions encoding beta-galactosidase-ATP/ADP translocase fusion
 RT proteins."
 RL Ann. N.Y. Acad. Sci. 590:397-407 (1990).
 RN [5]
 RP TOPOLOGY.
 RX MEDLINE=99141397; PubMed=9917392;
 RA Alexeyev M.F., Winkler H.H.;
 RT "Membrane topology of the Rickettsia prowazekii ATP/ADP translocase
 RT revealed by novel dual pro-lac reporters."
 RL J. Mol. Biol. 285:1503-1513 (1999).
 CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
 CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.
 CC THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF
 CC RICKETTSIAL PARASITISM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL: M28816; AAA26382.1; -.
DR EMBL: AJ235270; CAA14524.1; -.
DR PIR: J00026; J00026.
DR InterPro: IPR004667; ADP_ATP_car.
DR Pfam: PF03219; TLC; 1.
DR TIGRFAMs: TIGR00769; AAA; 1.
KW Transmembrane; Transport; ATP-binding; Multigene family;
KW Complete proteome.
FT DOMAIN 1 33 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 34 54 PROBABLE.
FT DOMAIN 55 67 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 68 88 PROBABLE.
FT DOMAIN 89 92 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 93 113 PROBABLE.
FT DOMAIN 114 147 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 148 168 PROBABLE.
FT DOMAIN 169 184 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 185 205 PROBABLE.
FT DOMAIN 206 218 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 219 239 PROBABLE.
FT DOMAIN 240 279 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 280 300 PROBABLE.
FT DOMAIN 301 320 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 321 341 PROBABLE.
FT DOMAIN 342 348 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 349 369 PROBABLE.
FT DOMAIN 370 379 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 380 400 PROBABLE.
FT DOMAIN 401 438 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 439 459 PROBABLE.
FT TRANSMEM 460 465 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 466 486 PROBABLE.
FT DOMAIN 487 498 CYTOPLASMIC (PROBABLE).
FT DISULFID 37 85 PROBABLE.
FT NP BIND 436 442 ATP (POTENTIAL).
SQ SEQUENCE 498 AA; 56821 MW; D3C450D2BC0BE53E CMC64;

Query Match 40.2%; Score 1056; DB 1; Length 498;
Best Local Similarity 41.6%; Pred. No. 1,1e-59;
Matches 207; Conservative 106; Mismatches 165; Indels 20; Gaps 7;

QY 2 TKTEKPFKGRSLPHTHEKLVLPMPFMCIFTFNYTVLRDPTDTLVGAPSGAE 61
DB 3 TSSESTLSELRKTIWPIBOYENKFLPLAFMFCILINSTLSIKDGFVV--TDIGTE 60
QY 62 AIPFKEFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPPLIFPFLFYIYPLRDVLT 121
DB 61 SISPLKTYIVLPSAVIMIIYVKCDILKQENVFVITSFLGFLAFVLYPDPVH 120
QY 122 PTEPADRLQALIPGGLGLVALILNNWTFAPFYVLAELGWSVMSLWFGFANETIKHEA 181
DB 121 PDHKTISLSLAVNPFKFIKIVGKWSFYTIAELGTWMLSLLEWQFANQITKIAEA 180
QY 182 KRFPALGIGNISILASGRVWASKLRASVSEGVDPWGSLSL---LMAAMTVGSLV 237
DB 181 KRFPMSFGCLANLALPVTSVIVIGFLHEKTIQI-----VAEHLKVPFLVIMITSFL 232
QY 238 LMASYMMINKNVLTDPRFYNEEMQKKGAKPKMNKMSFLYDRSPYILLTLVLVAY 297
DB 233 IILYLRMKNKVLTDPRLYDPAIVK--EKTAKALSFESLKMFTSKTYGVYAILLITAY 290
QY 298 GICINLIEVTWKSQKLQYPMNDYSEFGNFSFWTGVSVLMLFVGVGNIRKFGMLTG 357
DB 291 GVSVNLVEGVKWSKVELYPTKEAYTIYMGQFOFQGVVAIAFVNL-IGSNILRKVSWLTA 349
QY 358 ALVTPVWVLGLIYFPLVIFRNO-ASGLVAMFGTTPMLAVVGAIONITSKTKYALF 416
DB 350 AMITPLMNFITGAAPFSFIFDSVIANMLTGLASSPLTLLAVMGIMQNVLSKGVKSLF 409
QY 417 DSTEMAYIPLDQOKVYKGAADIVVAARFGKSGGALLIQGLVYICSGIGAM--TPYLAV 474
DB 410 DATGMAYITPLDKOLRVKGAAYEVYIGRLGKSGGALLIQSFIFLFPVGFIEATPYFAS 469

```

```

QY 475 ILLEFIAMLVSAFKLNK 492
DB 470 IFFIVILMTFAVKLNK 487

RESULT 8
TLC2_CHLN STANDARD; PRT; 540 AA.
ID TLC2_CHLN
AC Q927U0; Q9U010;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ADP ATP carrier protein 2 (ADP/ATP translocase 2).
GN TLCB OR ADT.2 OR CPN0614 OR CP0133.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_Taxid=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Iammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.",
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Burnham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Baas S.,
RA Linher K., Weidman J., Knouri H., Craven B., Bowman R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
RT pneumoniae AR39.",
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hatori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RL from Japan and CWL029 from USA.",
RN Nucleic Acids Res. 28:2311-2314(2000).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001646; AAD18753.1; -.
DR EMBL: AE002174; AAF38016.1; -.
DR EMBL: AP002547; BAA98821.1; -.
DR PIR: B72056; B72056.
DR PIR: C86567; C86567.
DR TIGR: CP0133; -.
DR InterPro: IPR004667; ADP_ATP_car.
DR Pfam: PF03219; TLC; 1.
DR TIGRFAMs: TIGR00769; AAA; 1.
KW Transmembrane; Transport; ATP-binding; Multigene family;
KW Complete proteome.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 94 114 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.

```

FT TRANSMEM 223 243 POTENTIAL.
 FT TRANSMEM 295 315 POTENTIAL.
 FT TRANSMEM 337 357 POTENTIAL.
 FT TRANSMEM 367 387 POTENTIAL.
 FT TRANSMEM 391 411 POTENTIAL.
 FT TRANSMEM 458 478 POTENTIAL.
 FT TRANSMEM 480 500 POTENTIAL.
 SQ SEQUENCE 540 AA; 60050 MW; 5746023AB02FABA3 CRC64;

Query Match 39.0%; Score 1025.5; DB 1; Length 540;
 Best Local Similarity 39.4%; Pred. No. 9.8e-58;
 Matches 214; Conservative 107; Mismatches 185; Indels 37; Gaps 9;

QY 1 MTKNEKPKRSLRSLPIHHEKLVPMFLMFCITENYIVADTDITLIVAGSGAEIPRIKF 60
 DB 1 MOSSVKKFSSRRALCPYKSEFSKVPFLFLAFVFGNCLLKNMKTIVIGSDXGA 60
 QY 61 EAIPIKFWLWPCAIIFMLIYAKLSNLSKOALFYAVGTPELFFALFPTIYPLRDVL 120
 DB 61 EVIPLKMGIVPGAVITVWYGMGSRYPBDTVFYCMAAFLGFFLFAVIIVYVGDSL 120
 QY 121 HPTFADRLQALIPGLIGLVAILENMTFAFYVLAELMGVMSLMTWGPANEITKTHE 180
 DB 121 HLNSLADKLOELPQGLRGFIWVRYSYIYVMSSELMSSVLSMLFWGLANOITITTE 180
 QY 181 AKRFALFEGIGANTSLASGRAIVASKLK-ASVSEGVDPW-GISRLMAMTIVSGIVL 238
 DB 181 AGRFALINTGLNLSICAGELISYMGKOTFYAISFACDSHVSMLNLTMLIT-CSGLIM 239
 QY 239 MASVWINKNV-LTDPFRYVPEEMQKKGK-----KPKNMKDSFLYDRSPYI 287
 DB 240 I-----MLVRRIHHLITDTSIPSRVLAEBGAATANLEKKKKPKAKANLFLHLQSYL 295
 QY 288 LLLTLVIAIGICINLIVTWKSQLQYPMNDYSEPMGNSFMTGVSVLIMLFGVGN 347
 DB 296 LGLIIVSYNLVHLFEVWVKDQVQSYSSHVEENGYSRITTLIGVSVLAVALTLGQ 355
 QY 348 VIRRGWLTGALNPPVWMLTGIVEFALVI-FRNOAGSLVMEFTPLMLAVVGAIONI 406
 DB 356 CIRKMGWVGLVLPVLMVLSGLFPFTPAKRDISIFGGVGLGTMPLAAMTGQWNV 415
 QY 407 LSKSTKVALPSTKEMAYIPLDOEKVKGKAAIDVVAARFGSGGALLIOGSLVIGSIG 466
 DB 416 LSRGKFFPQOTEMARIPSPBDKNKGKAAIDGVSRICKSGSLYQGLIVIFSSVA 475
 QY 467 AMPTVLAVILFIAIWLVSATKLNKFLAOSA-----LK-----DOEVAQE 508
 DB 476 ASLNVIALVLLIIMVWIAVAVIGKEYYSRAADAVATLKOPKEPSSSIVREAOESVQE 535
 QY 509 DSA 511
 DB 536 EWA 538

RESULT 9
 TLCC2_RICPR STANDARD; PRT; 501 AA.
 AC Q9ZD67;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP/ATP carrier protein 3 (ADP/ATP translocase 3).
 GN TLCC OR TLCC3 OR RP477.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
 OX NCBI_Taxid=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Magrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.B., Zomorodipour A., Andersson J.O.,
 Sichteritz-Ponten T., Alenmark U.C.M., Podowski R.M., Naeslund A.K.,

RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RL mitochondria".
 CC Nature 336:133-140(1998).
 CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
 CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.
 CC THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF
 CC RICKETTSIAL PARASITISM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ235271; CAA14932.1; -.
 DR PIR; B71707; B71707.
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC_1.
 DR TIGRFams; TIGR00769; AAA_1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 SQ SEQUENCE 501 AA; 57182 MW; D1EC540EB6D7E91F CRC64;

Query Match 36.9%; Score 969.5; DB 1; Length 501;
 Best Local Similarity 39.6%; Pred. No. 3e-54;
 Matches 197; Conservative 107; Mismatches 176; Indels 17; Gaps 7;

QY 9 FGKRSFLMPLHHEKLVPMFLMFCITENYIVADTDITLIVAGSGAEIPRIKF 68
 DB 8 FEKKEIIPERKEKELFLPMALMMLCLIFNFAKLSIDSLV--PSGAEIISFLKL 65
 QY 69 WLVPVCAIIFMLIYAKLSNLSKOALFYAVGTPELFFALFPTIYPLRDVLHPTFADRL 128
 DB 66 WLVPSCVITITLVKLSNKLNPYIYISVIGTFLFFLFAVIIVYNODIYHNDAMIN 125
 QY 129 LQALIPGLIGLVAILENMTFAFYVLAELMGVMSLMTWGPANEITKTHEAKRYALF 188
 DB 126 NLASVFNLMKFIKIGSKMSYALMYIPELSMAVINMLMFWQFNHIFDPAKARFYPVL 185
 QY 189 GIGANISLASGRAIVASKLKASVSEGV--DPR-----GISRLMAMTIVSGIVLMA 240
 DB 186 GMDGNIGLIIAGSVLVFSSGQYIIDEELLTDSYNSSNSNIMLQPIISIIIVAGIIMF 245
 QY 241 SYWINGNVLTDPFRYVPEEMQKKGKAKPKPMKDSFLYDRSPYILLTLVIAVIGIC 300
 DB 246 LFRILNKFILTNS--INVLQVKKVAATKTKLALIBESIKLIHSGYIGRIALLIICGLL 303
 QY 301 INLIEVTWKSQKLQYPMNDYSEPMGNSFMTGVSVLIMLFGVGNVIRKFGMLTGALV 360
 DB 304 INIVEGPMKAKIKELHENTVDYVNFMGFMNIMWGISCVTEMI-IGSNILRLGLISALL 362
 QY 361 TPVWVLLTGIVFPALVIFRNOAGSLVMEFTPLMLAVVGAIONIISKSTKVALPSTK 420
 DB 363 TPVMLSTGEMFPLFIIFIEIGTCGFENL--LYAIIIGVALONILSKSKSLPSTK 420
 QY 421 EMAYIPLDOEKVKGKAAIDVVAARFGSGGALLIOGSLV--CGSIGAMTPVLAVILF 478
 DB 421 EMAYIPLSLERKKGKAAVEVIGRKPSLGAFQSLIFITIPATDSIIITVLVFIIV 480
 QY 479 IIAIWLVSATKLNKFL 495
 DB 481 MNMLMWNIIKLNKEYI 497

RESULT 10
 TLCC2_CHLTR STANDARD; PRT; 540 AA.
 AC O64502;

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP/ATP carrier protein 2 (ADP/ATP translocase 2).
 GN TLCD OR CT495.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 NCBI_Taxid=813;
 [1]
 RN NCBI_Taxid=813;
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UM-3/Cx;
 RX MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,
 Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis."
 RL Science 282:754-759(1998).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE001323; AAC68096.1; -
 DR PIR; E71503; E71503.
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 FT TRANSMEM 23 43 POTENTIAL.
 FT TRANSMEM 61 81 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 150 170 POTENTIAL.
 FT TRANSMEM 185 205 POTENTIAL.
 FT TRANSMEM 222 242 POTENTIAL.
 FT TRANSMEM 292 312 POTENTIAL.
 FT TRANSMEM 334 354 POTENTIAL.
 FT TRANSMEM 361 381 POTENTIAL.
 FT TRANSMEM 389 409 POTENTIAL.
 FT TRANSMEM 455 475 POTENTIAL.
 FT TRANSMEM 477 497 POTENTIAL.
 SQ SEQUENCE 540 AA; 59764 MW; 126398B58C4A17AB CRC64;
 Query Match 36.4%; Score 956.5; DB 1; Length 540;
 Best Local Similarity 37.2%; Pred. No. 2.2e-53;
 Matches 197; Conservative 116; Mismatches 193; Indels 23; Gaps 7;
 QY 4 TEEKPFGRLRSFIMHITHELKVLPMFLMPCITFNNTVLRDPTDILVAGAPSGAAI 63
 DB 3 SEVSFSFSGYFPPIYSEFSKPLPLFLAFVGVNALLKTKYDLSLVGSRGAEVI 62
 QY 64 PFIKFWLVPCAIIFMLIYAKISNLSKQALFYAAGTFFLIFFALFPVIVPLRDVHPT 123
 DB 63 PFLKVGIVPGAVIVTMVIGMSRYSRGIVFISLVGFLGFPALFAVIVIPIDGALHN 122
 QY 124 EPADRLQAILPPLGLGLVAIRNMTFAAPVYLAEIWSVMLSLMFWGFANETIKHEKR 183
 DB 123 KLAAKLOSILPPGGGFVMTQVWSYSLSYVMSISIVSTLFWGVANHITSVREAGR 182
 QY 184 FYALFGGANISLASGRAIYMASKLRA-SVSEGVDPK-GISRLMAMTVISGLVYMAS 241
 DB 183 FYALINTGLANSSVPAGEVSIMLGRNVIAPMAVDPHHEVLNLTLLIVLAGVTL- 239
 QY 242 YWMINKVNLTPRFYNP-EMOKGKGAKPKMMKDSFLYIDRSPLYLLTL 292
 DB 240 YLYGKGLDRIMDETSMLEKGLAEMSVAGLTKKEKRSKAK-AKSLFALLLSRYLGLAV 297

QY 293 LVIAVICINLIEYTKSOLKQYPMNDYSEFMGNFSFWGVSVLIMLFGVGNIRKF 352
 DB 298 VVLSYNLVHLFEVWMDQCVRIYASRVERFSYMSRITTLTGIVSALGIAAGQTLNRW 357
 QY 353 GMLTGAIVTPVWVLLTGVIF-ALVIFPNQASGLVAMFGTTPVLMVAVGAIONLSKST 411
 DB 358 GWITGALVPLTMLITGALFFGALYAVKGDAMITGILGISPLVITMLGSGVQVFSRAI 417
 QY 412 KYALPDSIKEMAYIPLDQKVKKKAIDVVAAPFGSGGALLIQGLIVIGSIGAMTPY 471
 DB 418 KFTYFDQTKEMAFIPLEDEKNGYKKAIDGVISRVKSGSGSLVYQGLIIFSSVAASINA 477
 QY 472 LAVLLPFIATWVSAKTKLFLAQA-----LKQEVNQEBSAPAS 514
 DB 478 ITIVLLALGSMIVIAMLGREYAKTEALFRVNVSESDVQERREAS 526
 RESULT 11
 ID TLCD CHLMU STANDARD, PRT, 543 AA.
 AC Q9PJP6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE ADP/ATP carrier protein 2 (ADP/ATP translocase 2).
 GN TLCD OR TC0782.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 NCBI_Taxid=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MOPN / N199;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gyll S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uckerback T., Berry K., Baas S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwynn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
 RT pneumoniae AR39."
 RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE002346; AAP39585.1; -
 DR PIR; F81665; F81665.
 DR TIGR; TC0782; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 FT TRANSMEM 23 43 POTENTIAL.
 FT TRANSMEM 61 81 POTENTIAL.
 FT TRANSMEM 92 112 POTENTIAL.
 FT TRANSMEM 150 170 POTENTIAL.
 FT TRANSMEM 185 205 POTENTIAL.
 FT TRANSMEM 222 242 POTENTIAL.
 FT TRANSMEM 292 312 POTENTIAL.
 FT TRANSMEM 334 354 POTENTIAL.
 FT TRANSMEM 361 381 POTENTIAL.
 FT TRANSMEM 389 409 POTENTIAL.
 FT TRANSMEM 452 472 POTENTIAL.
 FT TRANSMEM 477 497 POTENTIAL.

SQ SEQUENCE 543 AA; 59672 MW; 6291A671F1924D7D CRC64;
 Query Match 36.2%; Score 953; DB 1; Length 543;
 Best Local Similarity 37.5%; Pred. No. 3.6e-53;
 Matches 199; Conservative 11; Mismatches 197; Indels 24; Gaps 7.

4 TEEKFGKLRSLRMPHITHELKVLPMFLMPCITFNVTVDLDTOTLIIVGAPGSGAAI 63
 3 SEVKTFKSRFRKFFPHKSEFPKFFPLLLAFVGVGNISLKTITDLSVLASRGAEVI 62
 64 PEIKFVLVPCAIIFMLIYAKLSNLSKQALFYAVGTPPELLIFPALPPTVYPLRDVLHT 123
 63 PFLKVMGIVGAVITIMTIYGMMSCRYSRGFVCAIVGGLSPFALFACIYPMGALHIN 122
 124 EPADLQALIPPGILGLVALIRNTPFAFYVLAELMGSMLSMFGANETTKIHEAKR 183
 123 GLAALQTLIPRGAGFVVMQVMSYSLYVMSLSMSSVSLTFLWGALNHITSVREAGR 182
 184 FYALFGIGANISILASGRAIVASK-LRASVEGVDPV-GISRLMLMTIVSGLVMAS 241
 183 FYALINVGAINVSIYAGEISLMLGKHTLIPSSMAVDAMHGVLLNITLIIVAAGLIL- 239
 242 YWINKNVLTDP-----RFYNPEWQKGGKAKPRMNMKOSPLYLDRSPYILLTL 292
 240 YLYRLDLHTEAPVLGDVSEMSVAQLKQEK--RPKAKXSLSLVLFRRYLMGIAY 297
 293 LVIANGICINLIEVWKSQLQYPMNDYSEFMNGSEFMGVSVLMLFVGNVIRKF 352
 298 VLAALNLAHLELVVMSKQVCOIYSSRVEFNSYMRITAFITGIVSALAGVFAAGSIRRW 357
 353 GMLTALVTPVWVLLTGIVFF-ALVIFRNQASGLVAMFGTPPLMAVVGAIQNLISKST 411
 358 GNTVALITLPLTMLITGLIFGAIYAVKGDAMITLGLFESPLVITANLGVQNVFSAL 417
 412 KYALFDSIKEMAYIPLEDOQVKGAAIDVVAARFGSGGALLIOGLVIGSGIGAMTPY 471
 418 KFYEDQTKEMAFIPLDEDEKQYKAAIDGVISRVKSGSILYQALLIFESVADCNNA 477
 472 LAVILLFIATMVLSATKNTKFLAQS-----ALKQEVAAQEDSARASS 515
 478 ITIVILLALGGMVWVWVWLKREYSVTALGKARAEPSPLODEDESRVSS 528

RESULT 12
 TLCD_RICPR STANDARD; PRT; 512 AA.
 AC Q9ZD47;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP carrier protein 4 (ADP/ATP translocase 4).
 GN TLCD OR TLCD4 OR RP500.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Anderson S.G.E., Zomrodipour A., Andersson J.O.,
 RA Sichenitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria."
 RL Nature 396:133-140(1998).
 CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
 CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.
 CC THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF
 CC RICKETTSIAL PARASITISM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL; AJ235272; CAA14952.1; -
 DR PIR; F71653; F71653.
 DR InterPro; IPR004667; ADP-ATP-car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 SQ SEQUENCE 512 AA; 58208 MW; D09A265DC38936EB CRC64;

Query Match 33.5%; Score 880.5; DB 1; Length 512;
 Best Local Similarity 36.9%; Pred. No. 1.3e-48;
 Matches 184; Conservative 108; Mismatches 177; Indels 29; Gaps 10;

9 FGKLRSLRMPHITHELKVLPMFLMPCITFNVTVDLDTOTLIIVGAPGSGAAIPIPKF 68
 19 FSKLTVDYIMPIKRHEISKFLPITLMFCILFIQMLIRALKDSIVTWM-IGAEITSPKLF 76
 69 MLVVPICALIIFMLIYAKLSNLSKQALFYAVGTPPELLIFPALPPTVYPLRDVLH--PTEFA 126
 77 GGVNPSAFILVITVYKLVNRRKANITFYLLISITLTFALFAYVIFPHMLHLPVT-V 135
 127 DRLQALIPPGILGLVALIRNTPFAFYVLAELMGSMLSMFGANETTKIHEAKRYA 186
 136 HNLTRSL-PNLKWFITLLSKMSFSLFYIABLWNVVAFALLFMQFVNNTTVESSKPYF 194
 187 LFGIGANISILASGRAIVASKLRASVSEGVDPG---ISRLMLMTIVSGLVMAASY 242
 195 LFGILSGTGYLACHFLENLSNIVYVYTNKRALOSSPHRTSIQIILTVLIGVISIKTF 254
 243 WINKNVLTDP-----RFYNPEWQKGGKAKPRMNMKOSPLYLDRSPYILLTLVIAY 297
 255 WLNHAKVLDKKNMALRF-----KTKNSKITAKSFQMLISRIRLITLILCY 304
 298 GICINLIEVWKSQLQYPMNDYSEFMNGSEFMGVSVLMLFVGNVIRKFGMLTG 357
 305 GIALLVEGPKAAATITKPTTEYAFISGLYSTGVFTIFVL-LGSNIVRMGMETS 363
 358 ALVTPVWVLLTGIVFFALVIFRNQASGLVAMF-GTTPMLMAVVGAIQNLISKSTKALF 416
 364 AVITPSIVFITGLIFAVNNFEFGAGLIANFILTDPALVAITGAIQNVLSKSKYTLF 423
 417 DSTEMAYIPLEDOQVKGAAIDVVAARFGSGGALLIOGLVIGSGIGAMTPYAVIL 476
 424 DSTEMAYIPLEPIKISGRAADVITGKLGSGSAPLQSLIFITLPSASQISISICLMI 483
 477 LFIIF-AIWLVSATKLNK 492
 484 IFILTCTWIMATKELNK 501

RESULT 13
 TLCE_RICPR STANDARD; PRT; 500 AA.
 ID TLCE_RICPR
 AC O05962;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP carrier protein 5 (ADP/ATP translocase 5).
 GN TLCE OR TLCS OR RP739.
 OS Rickettsia prowazekii.
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=Madrid E; PubMed=9823893;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
 RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RL Nature 396:133-140(1998).
 RN [2]
 RP SEQUENCE OF 325-500 FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=97419517; PubMed=9274032;
 RA Andersson J.O., Andersson S.G.E.;
 RT "Genomic rearrangements during evolution of the obligate
 RT intracellular parasite Rickettsia prowazekii as inferred from an
 RT analysis of 52015 bp nucleotide sequence.";
 RL Microbiology 143:2783-2795(1997).
 CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
 CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.
 CC THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF
 CC RICKETTSIAL PARASITISM (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AJ235273; CA15167.1; -
 DR EMBL, Y11778; CA172457.1; -
 DR PIR, G71633; G71633.
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam: PF03219; TLC; 1.
 DR TIGRFAms: TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 FT TRANSMEM 26 46 POTENTIAL.
 FT TRANSMEM 62 82 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 224 244 POTENTIAL.
 FT TRANSMEM 287 307 POTENTIAL.
 FT TRANSMEM 328 348 POTENTIAL.
 FT TRANSMEM 357 377 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 469 489 POTENTIAL.
 SQ SEQUENCE 500 AA; 57073 MW; F83DB48D08CF5F72 CRC64;
 Query Match 33.9%; Score 865; DB 1; Length 500;
 Best local similarity 35.7%; Pred. No. 1,2e-47;
 Matches 176; Conservative 109; Mismatches 182; Indels 26; Gaps 10;
 QY 11 KLSFLPPIHTEHLEKVLPMFLMPCITFNTYVLRDVKTLIVGAPSGAALPFIKFWL 70
 DB 13 KFRALFVPHVNEELGKPIPLSMFLIFNQNIRILKDSILISE--ISAEIAGFAKYVC 70
 QY 71 VVPAALIFEMLIYALNLSKQALFVAVGTFFLFPALFVTVPLADV--LHTEPADR 128
 DB 71 VTPVAAAFVILIYAMINHLTFEKFYUSAFIFSCFLIFAVIIPNHIHFVHDITLSDW 130
 QY 129 LQALILPGLGLVAILRMWTFPAAFVYLAELMGVSLIMFGFANEITKIEAKRYALF 188
 DB 131 MNKY--EHEKRYILVNGWGIYVYSLAELMPNIFVYLFFQFNEILTTEARQFTLP 188
 QY 189 GIGANISLASGRALVWASKLRASVSEGVDPWGISLRLLMAMTVSGLV-----LMASYW 243
 DB 189 SLFNSSLIIVGFLLMNLISSEDTIIKKPISSDSKITLVQSTIIAIVAILCCILVRF- 247

QY 244 WINKNVLTDRPVPNEEMOKGKGAEPKMMKDSFLYLBSPYLLTLIVAGICINL 303
 DB 248 -ISKYITNLPFH--KTSSTSTAPRMGLISFKYIVSKYIMLLIGSAFPAINL 303
 QY 304 IEVWKSQQLQYPPNNNDYSEFMGNFSGMTGVSVLLIMLFPVGNVIRKFGMLGALVTPV 363
 DB 304 VEAVWAKAKIKELVETVTVTAEPNSFLYLTWG-VAIIMVTIIGNNVMEHMFVAIVSPV 362
 QY 364 MVLITGVFPALVIFRNOA---SGLVAMFGTTPMLAVVGAIONLSTKALDST 419
 DB 363 IIMVGTLPFGLVFDDQIISLFGDALIM--SPLAAVSIIGGIQNIILAKGTYSIDTS 419
 QY 420 KEMAYIPLDQOKYKGAALDVVAARFGSGGALIQGL--LVYICSGIAMTPYLAIVIL 477
 DB 420 REMIYPLDDELTKTGAADVISAQVKGSSGLVQSIIFTLVNAFTSISPLIMVVT 479
 QY 478 FIAIMVLVSATKL 490
 DB 480 FVCPAMIVAVRKI 492
 RESULT 14
 ID TLCB RICPR STANDARD; PRT; 507 AA.
 AC 09ZDF2;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP/ATP carrier protein 2 (ADP/ATP translocase 2).
 GN TLCB OR TLC2 OR Rp377.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 ON NCBI Taxid=782;
 RX SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
 RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RL Nature 396:133-140(1998).
 CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
 CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.
 CC THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF
 CC RICKETTSIAL PARASITISM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AJ235271; CA14836.1; -
 DR PIR, B71695; B71695.
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam: PF03219; TLC; 1.
 DR TIGRFAms: TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 SQ SEQUENCE 507 AA; 58457 MW; 00DB2DFCC51D9D59 CRC64;
 Query Match 31.3%; Score 824; DB 1; Length 507;
 Best local similarity 34.2%; Pred. No. 4.5e-45;
 Matches 175; Conservative 114; Mismatches 170; Indels 52; Gaps 10;
 QY 11 KLSFLPPIHTEHLEKVLPMFLMPCITFNTYVLRDVKTLIVGAPSGAALPFIKFWL 70

```

Db 18 KFRIVWPVRSYELTKFIPMLMFILLNINVRISDSVVLISS--EVLSEFIKMG 75
Qy 71 VVPCALIFMLIYAKLSNLSKQALFYAVGTPPLIFPALFPVIVYPLRDVLPTEPADRLQ 130
Db 76 EMPGVFVLIYSKLCNIMTEQVFRITSTFFLFFALFGLFPYKEFFHNNELINOY 135
Qy 131 AILPGLGLGVAAILRNWTFPAFVLAELMGSMVLSIMWGANETTKTKEKREYALFGI 190
Db 136 IIVPHLKWFLIIMGQSVLFTYIMGELMPVIVFLWQJLANKITVKEBAPREYSEFTL 195
Qy 191 GANISLASGRAIYVASK-----LRASVEGVDPWGISRLMAMTIYSGVLVMSYW 243
Db 196 FQGNILFSGVLIYIFPAKSEHFLPLFAHNL---DTNILLKSFITVILISGLCLMLHK 252
Qy 244 WINNVLTDPFVYNEEMQKKGAKP-----KANKDSFLYLDSPYILLTLVAYG 298
Db 253 LIDKSVV-----EADKNIKFKRNORTDILKLSLSAKIILTSRLGFCILLVMSYS 303
Qy 299 ICILILEVTWKSOLQVPMNDYSEPMGNFSFMTGVSVLIMLVGNGVIRKFGMLTGA 358
Db 304 MSNILEGLMWSKVQKLPATKDFISTHGEVLFWTGVLT-LVSAFLSGSLIRITGMWGA 362
Qy 359 LVPVWVLLTGIVFPAVIFRNQASGLVAMFG-TTPLMLAVVGAIONILSKSTKALFD 417
Db 363 IITPIMFVAVGMFFSFTIFEGHLGNIVNTLGYSPLVIVIFIGLWVFAKXKSLFD 422
Qy 418 STKEMAYIPLDOEKVKGKALIDVVAARFGSGGALIO-----QGLVYC 462
Db 423 ATKEMVYIPLDNEIKTKGKAADVWGAKIGKSGIAGIIOFISFIFPNVHNDIAGLL--- 479
Qy 463 GSIGAMTPYLAIVLLEFIATVLSATKLNKL 493
Db 480 -----WTFIIVCILMLYGVKVL-SQYTKM 504

```

```

RESULT 15
LACY LEULA
ID LACY LEULA STANDARD; PRT; 639 AA.
AC Q4624;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactose permease (lactose-proton symport) (Lactose transport protein).
GN LACS.
OS Leuconostoc lactis.
OG Plasmid pNZ63.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_Taxid=1246;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NZ6009;
RX MEDLINE=96209221; PubMed=8633855;
RA Vaughan E.E., David S., de Vos W.M.;
RT "The lactose transporter in Leuconostoc lactis is a new member of the
RT Lacs subfamily of galactoside-pentose-hexuronide translocators.";
RL Appl. Environ. Microbiol. 62:1574-1582(1996)
CC -1- FUNCTION: RESPONSIBLE FOR TRANSPORT OF BETA-GALACTOSIDES INTO THE
CC CELL, WITH THE CONCOMITANT EXPORT OF A PROTON (SYMPORT SYSTEM),
CC AND ALSO FOR TRANSPORT OF HOMOLOGOUS AND HETEROLOGOUS EXCHANGE OF
CC BETA-GALACTOSIDES.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- DOMAIN: THE ENZYME IIA-LIKE REGION MAY SERVE A REGULATORY
CC FUNCTION.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC SODIUM-GALACTOSIDE SYMPORTER FAMILY (SGF).
CC -1- SIMILARITY: Contains 1 PTS EIIA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

entitles requires a license agreement (See <http://www.ebi-sib.ch/announce/> or send an email to license@ebi-sib.ch).

```

CC -----
CC EMBL; U47655; AAC44113.1; -.
CC HSSP; P45618; ZGPR.
CC InterPro; IPR001927; Na/Gal symport.
CC InterPro; IPR001127; PTS EIIA.
CC Pfam; PF00358; PTS_EIIA_1; 1.
CC ProDom; PD002243; PTS_EIIA; 1.
CC TIGRFAMs; TIGR00792; gph; 1.
CC TIGRFAMs; TIGR00830; PTBA; 1.
CC PROSITE; PS00371; PTS_EIIA_1; 1.
CC PROSITE; PS00872; NA_GALACTOSIDE_SYMPT; 1.
CC K1 Transmembrane; Sugar_cransport; Transport; Symport; Phosphorylation;
CC plasmid.
CC K1 DOMAIN.
CC FT 1 473 SGF DOMAIN.
CC FT 474 639 EIIA DOMAIN.
CC FT MOD_RES 557 557 PHOSPHORYLATION (BY HPR) (BY SIMILARITY).
CC FT TRANSMEM 20 40 POTENTIAL.
CC FT TRANSMEM 59 79 POTENTIAL.
CC FT TRANSMEM 99 119 POTENTIAL.
CC FT TRANSMEM 124 144 POTENTIAL.
CC FT TRANSMEM 176 196 POTENTIAL.
CC FT TRANSMEM 207 227 POTENTIAL.
CC FT TRANSMEM 264 284 POTENTIAL.
CC FT TRANSMEM 294 314 POTENTIAL.
CC FT TRANSMEM 323 343 POTENTIAL.
CC FT TRANSMEM 347 367 POTENTIAL.
CC FT TRANSMEM 398 418 POTENTIAL.
CC FT TRANSMEM 433 453 POTENTIAL.
CC SQ SEQUENCE 639 AA; 70153 MW; 25DF2819761B415 CRC64;

```

Query Match 5.2%; Score 135.5; DB 1; Length 639;
Best Local Similarity 20.0%; Pred. No. 0.13; 209; Indels 129; Gaps 22;
Matches 104; Conservative 78; Mismatches 209; Indels 129; Gaps 22;

```

Qy 54 GAPSGAEAIPIFIKFWLVPICAIIFMLIYAKLSNLSKQALFYAVGTPPLIFPALFPPTVI 113
Db 49 GLPQSVANKIKGLITLVAIVIRLAEVIDPILGNIVNTNRNGKPEWVIGAVSVS- 107
Qy 114 YPLRDVLPTEPADRLCALIIPGLGLGVAILRNW-----TPAAFYVLAELMGSMVLSMF 168
Db 108 -----LLVVFITGIFGLAHI--WIAIPAVFTVLFILIDIFYS-FADVAY 149
Qy 169 WGFANEITKIHAKRFALFG-----IGAN-----ISLASGRAIYVASKLRA 211
Db 150 WGVVPALISDSKSGKGIPTSGSFTGSGIWMGLTIVVPTTYFTFIATGK----- 200
Qy 212 SVSEGVDPW-GISRLMAMTIYSGVLVMSYWMINKVLTDPFVYNEEMQKKGAKP 270
Db 201 --EGPSPGWFQFSL-VVSIYAVSALAV-----AFGTREKONIRNATK 242
Qy 271 KANKDSFLYLDSPYILLTLVIAVGCINLIE--VTWKSQKLOYPMNNDYSEPMGN 328
Db 243 KTSIKDVFSGIINHDIQILMSLAVYVTVNGVLFEEFKVGLGKPN----- 292
Qy 329 FSPFTGVSVLIMLVFGVGNVIRKGMVLTGALVTPVWMLLGIYFPAVIFRNQASGLVAM 388
Db 293 -EFV-----IAGATVYIGFSTAPLVYLNKFTIRKVFISGQAMILSYLFTI 340
Qy 389 FGTPMLAVVGAIONILSKSTKALFDSTKEMAYIPLDOEKVKG-----KAIDVVA 443
Db 341 FGTNNMM--VTIGILNFNFAGLVVLSLDSIEYQGLKNGRNEAVVLAVRPMLDKIT 399
Qy 444 ARF--GKSGALLIQGLL--VICSGIGA-----MTPYLAIVLFIATVLSAT 488
Db 400 GAFSGNGLVGAIAITAGMTGATAGDISASKINTFEIYAFYTPPLIFSLIALVIFLMKVKIT 459
Qy 489 -----KLNKLFIAQSALK--EDEV---AQEDSAPAS 514
Db 460 EKHRAEIVIELKT-LSSGAKKANTSEVNVLEBEIFAPAS 498

```

```

RESULT 16
YH06_HAEIN STANDARD; PRT; 669 AA.
ID YH06_HAEIN
AC P45335;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable transport protein H11706.
GN H11706.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrman J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD."
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -1- SIMILARITY: BELONGS TO THE BCCT (TC 2.A.15) FAMILY OF
CC TRANSPORTERS.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U3843; AAC23352.1; -.
DR PIR: D64137; D64137.
DR TIGR: H11706; -.
DR InterPro: IPR000060; BCCT_transporter.
DR Pfam: PF02028; BCCT_1.
DR ProDom: PD010111; BCCT_transporter; 1.
DR TIGRFAMs: TIGR00842; bcct; 1.
DR PROSITE: PS01303; BCCT_1.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 48 68 POTENTIAL.
FT TRANSMEM 87 107 POTENTIAL.
FT TRANSMEM 139 159 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT TRANSMEM 224 244 POTENTIAL.
FT TRANSMEM 259 279 POTENTIAL.
FT TRANSMEM 314 334 POTENTIAL.
FT TRANSMEM 345 365 POTENTIAL.
FT TRANSMEM 397 417 POTENTIAL.
FT TRANSMEM 444 464 POTENTIAL.
FT TRANSMEM 470 490 POTENTIAL.
SQ SEQUENCE 669 AA; 75020 MW; 0AE8E803BB5C733F CRC64;
Query March 5.1%; Score 133; DB 1; Length 669;
Best Local Similarity 19.8%; Pred. NO. 0.19; Indels 216; Gaps 30;
Matches 116; Conservative 87; Mismatches 167;
15 FLNPIHTEHL-----KVLPMFLMFCITFNYTVLRDVKTLIAGAGSGAELPIFKFW 69
27 FIAEGQALINQAKSGIFANFSFWYVLTFSV-----FLGFL 63

```

```

QY 70 LVVPCALIFMLIYAKLSNI-----LSKALFPAVGTPF-LIFPALPPTVTP 115
DB 64 LI-----LSVSSIGNITLGDDEEPPEPFSLSMLMLPAAGMGVGFVGAEPITTHY 115
QY 116 LRDVLPTEPADLQAILPPGLIGLVAIIRNMTFAAFVYLAELGSMVSLMFWGANET 175
DB 116 LSDITGSAHKKQGBALL-----HTLFPHGIGHMAV--YGTALALAVGFPRYKL 163
QY 176 T-----KHEAKRYA---LFGI-----GANISLASG-RAIWAASKL 209
DB 164 PLALRSCFPYLKDRINGKIGDAIDWALATATFGIITTLGFPSSQIGALBOIGWISON 223
QY 210 RASVSGVDPMWGISLHLMAMTIVSGVLMASVWYINKVLTDRFPNPEMOKGKGAK 269
DB 224 SPALQGVLT--VVMCLAVFSAISGV-----GKGVK 252
QY 270 --PRMNMKDSF---LYLDRSPYLLTLTLVIAVG-ICIMLEIYVWKSQKLQYPMNDY 322
DB 253 ILSEINLTALFCLLPVLSGPTLYLSAFSDNIGNVFSVLVQLSFT-----YAYQEH 307
QY 323 SEFMGNFSF-----WGVSVLIMLVGVGNVIRKFMVLTGALVTPVWLLTGIVFPA 374
DB 308 TSWFSGWTVLYMAMWCSWAFVGLFARIKSGRTIREF--IFGLVLP---SLFGILMF- 361
QY 375 LVIFRNQASGLVAMFGTTPML--AVVGAIQWILSKSTKALPDSIKEMAYIPL----- 427
DB 362 -----TVFGNTAVLNDGIAAGLGERTIS-SPILLF--KFLNLYLPPTITG 405
QY 428 -----DOE-----QKVGKRAIDVVARFEGSGG-A 452
DB 406 FVSLVILVLFRTSADSGIVLVNNIASRDKSLASPAKQALIMKGTMSVVAIVLMQSGLA 465
QY 453 LIQGLIVIGSGIGAMTPYLAIVLIFTAIW--LVS-----ATKLN 491
DB 466 NLQMTMLIVA-----LPFALLMLVMCFSLMKLADKKYFSTRVN 505
RESULT 17
NARK_ECOLI STANDARD; PRT; 463 AA.
ID NARK_ECOLI
AC P10903;
DT 01-JUN-1989 (Rel. 11, Created)
DT 01-JUN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrite extrusion protein 1 (Nitrite facilitator 1).
GN NARK OR B1223.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN
RP SEQUENCE FROM N.A.
RP STRAIN=K12;
RX MEDLINE=89338707; PubMed=2668029;
RA Koji S., Nohno T., Saito T., Taniguchi S.;
RT "The nark gene product participates in nitrate transport induced in
RL Escherichia coli nitrate-respiring cells."
RL FEBS Lett. 252:139-143(1989).
[2]
RN
RP SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1245-1247(1997).
[3]
RN
RP SEQUENCE FROM N.A.
RP STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;

```

RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA itemo K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horuchi T.,
RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=9251357; PubMed=9097039;
RA Nobuo T., Noji S., Taniguchi S., Saito T.;
RT "The *narX* and *narX* genes encoding the nitrate-sensing regulators of
RT *Escherichia coli* are homologous to a family of prokaryotic two-
RT component regulatory genes.";
RL Nucleic Acids Res. 17:2947-2957(1989).
CC -1- FUNCTION: INVOLVED IN EXCRETION OF NITRITE PRODUCED BY THE
CC DISIMILATORY REDUCTION OF NITRATE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (probable).
CC -1- INDUCTION: BY NITRATE UNDER ANAEROBIC CONDITIONS.
CC -1- MISCELLANEOUS: SINCE THE TRANSPORT IS CONTROLLED BY OXYGEN, FIVE
CC CYS RESIDUES IN *NARX* MAY PLAY AN IMPORTANT ROLE IN SENSING REDOX
CC POTENTIAL TO REGULATE THE FACILITATOR.
CC -1- SIMILARITY: BELONGS TO THE *NARX*/NARX FAMILY OF TRANSPORTERS.
CC -1- SIMILARITY: STRONG, TO E. COLI *NARX*.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15996; CAA34126.1; -
DR EMBL: X69189; CAA48933.1; -
DR EMBL: AB000220; AAC74307.1; -
DR EMBL: D90757; BAA36091.1; -
DR EMBL: D90758; BAA36093.1; -
DR EMBL: X13360; CAA31740.1; -
DR PIR: S05239; GRECONK.
DR Ecogen, EG10642; *narX*.
DR InterPro: IPR004737; NO3 transporter.
DR TrEMBL: TIGR00886; 2A0108; 1.
KW Nitrate assimilation; Transport; Transmembrane; Inner membrane;
KW Multigene family; Complete proteome.
FT TRANSMEM 38 58 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 132 152 POTENTIAL.
FT TRANSMEM 178 198 POTENTIAL.
FT TRANSMEM 212 232 POTENTIAL.
FT TRANSMEM 257 277 POTENTIAL.
FT TRANSMEM 284 304 POTENTIAL.
FT TRANSMEM 320 340 POTENTIAL.
FT TRANSMEM 345 365 POTENTIAL.
FT TRANSMEM 405 425 POTENTIAL.
FT TRANSMEM 436 456 POTENTIAL.
SQ SEQUENCE 463 AA; 49693 MM; AF1D67CAD40FE8A CRC64;
Query Match 5.0%; Score 132.5; DB 1; Length 463;
Best Local Similarity 22.3%; Pred. No. 0.14;
Matches 108; Conservative 69; Mismatches 174; Indels 133; Gaps 28;
QY 69 MIVVPCALIMLIMLAKLSNI-LSKQALFYANGTFFLLFPYIVYPLDVLHPTFEAD 127
DB 39 WISVPCALLAFQWMLFSAAVAVNPKVGFNTTQLEMLRALPSVSGALRV--PYSF-- 94
QY 128 RLQALIPGLGLGVAIL--RWTFPAFYVLAELMGVSLSMFPGFANEITKIHAEAFY 185

DB 95 -----MWPIFGRRWT--AFST-----GLIILPCVWLGF-----VQDISTPY 130
QY 186 ALRGIGANISLLASGRAIVASKLRSVS-----EGVDPGIGISRLIMAM 230
DB 131 SVFII--ISLLGFAAGANASSM-ANISFFPRKQGGALGLNGGNGSVMOVLAP 186
QY 231 TIYSGVLMAVYMWINGVLTDPFRFYPEEMQKKGAKPKMMKOSFLYLDSPYLL- 289
DB 187 LVYS-LSIFAVF-----GSGQVKPDGTE--LVLAANSWITWP 221
QY 290 -LTLVIAVIGICINLIEVTKSQLQYPMND-----YSEKGNF-SFWTGV-- 335
DB 222 FLAIFITIAWFGWMDL-ATSKASIKELPVLKRGHLIMSLYLATPGSIFGSAGPML 280
QY 336 -----VSLIMLPFG--GNVIRKRW-----LNGALVTPV-----MVLITGVFPAV 376
DB 281 SKTQFPDVQILOYAFEPFPIGALARSAGALSRLGTGRTVLVFLIMAFSGLLFTLP 340
QY 377 IFRNQSGLVAMFGTFPLMAVVV-----GAIONILSKSTKXALPSTKEMAVIPDOE 430
DB 341 T-DQGGSPMAFPA--VFALFLTAGLGGSGTFQMISVIFKLTMDRVAAEG--GSD 392
QY 431 QYKGAALDVAA-----RFGSGGALIQGL--LVIGGS-IGAMTPYLAIVLFI 481
DB 393 FRAMREKATDTAALGFIASIGAIIGFPIKAFGSSIALTGSVGMKVLIFIACVI 452
QY 482 TWLV 485
DB 453 TWAV 456

RESULT 18
ID NARU_ECOLI STANDARD; PRT; 462 AA.
AC P37758; P77696;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrite extrusion protein 2 (Nitrite facilitator 2).
GN NARU OR B1469.
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Bonney V., Ratouchniak J., Biasco F., Chippaux M.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.,
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=9251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kaai H., Kashimoto K., Kimura S., Kitagawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sempel G., Seki Y., Sivasubraman S.,
RA Tagami H., Takeda U., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horuchi T.;
RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).

[4]
 RP SEQUENCE OF 410-462 FROM N.A.
 RX MEDLINE=91042410; PubMed=2233673;
 RA Blasco F., Jobbi C., Ratouchniak J., Bonnefoy V., Chippaux M.;
 RT "Nitrate reductases of *Escherichia coli*: sequence of the second
 RT nitrate reductase and comparison with that encoded by the *nargH1*
 RT operon.";
 RL Mol. Genet. 222:104-111(1990).
 RN [5]
 RP IDENTIFICATION.
 RX MEDLINE=95266795; PubMed=7747940;
 RA Bonnefoy V., Demoss J.A.;
 RT "Nitrate reductases in *Escherichia coli*:";
 RL Antonie Van Leeuwenhoek 66:47-56(1994).
 CC -1- FUNCTION: INVOLVED IN EXCRETION OF NITRATE PRODUCED BY THE
 CC DISIMILATORY REDUCTION OF NITRATE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Probable).
 CC -1- MISCELLANEOUS: SINCE THE TRANSPORT IS CONTROLLED BY OXYGEN, FIVE
 CC CYCLES RESIDUES IN NARU MAY PLAY AN IMPORTANT ROLE IN SENSING REDOX
 CC POTENTIAL TO REGULATE THE FACILITATOR.
 CC -1- SIMILARITY: BELONGS TO THE NARX/NASA FAMILY OF TRANSPORTERS.
 CC -1- SIMILARITY: STRONG, TO E. COLI NARX.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X94992; CA64448.1; -
 DR EMBL: AB000244; AAD13433.1; -
 DR EMBL: D90788; BA15120.1; -
 DR EMBL: D90787; BA15118.1; -
 DR EMBL: X17110; -; NOT_ANNOTATED_CDS.
 DR PIR: H64899; S11431.
 DR EcGene; EG12153; narU.
 DR InterPro; IPR004737; NO3_transporter.
 DR TIGRFBMs; TIGR00886; 2A0108; 1.
 KM Nitrate assimilation; Transport; Transmembrane; Inner membrane;
 KM Multigene family; Complete proteome.
 FT TRANSMEM 36
 FT TRANSMEM 77
 FT TRANSMEM 102
 FT TRANSMEM 126
 FT TRANSMEM 146
 FT TRANSMEM 181
 FT TRANSMEM 201
 FT TRANSMEM 207
 FT TRANSMEM 227
 FT TRANSMEM 259
 FT TRANSMEM 279
 FT TRANSMEM 288
 FT TRANSMEM 308
 FT TRANSMEM 318
 FT TRANSMEM 338
 FT TRANSMEM 345
 FT TRANSMEM 365
 FT TRANSMEM 402
 FT TRANSMEM 422
 FT TRANSMEM 433
 FT TRANSMEM 453
 FT TRANSMEM 290
 FT TRANSMEM 293
 FT TRANSMEM 293
 FT TRANSMEM 462 AA; 49890 MW; 335B3C140FBC4DF CRC64;
 SQ SEQUENCE

Query March 4.9%; Score 129; DB 1; Length 462;
 Best Local Similarity 22.3%; Pred. No. 0.24; Mismatches 171; Indels 134; Gaps 26;
 Matches 108; Conservative 71; Mismatches 171; Indels 134; Gaps 26;
 QY 69 WLVPVCAIIFMLIYAKLSNT--ISKQALFYAVGTPPLIFALFPTV-----LY 114
 DB 37 WISVSCILLAFVCMWLSAVTVNINKIGENFTDQLFL--TALPSVSGALLRVPSFMV 94
 QY 115 PLRVVLPTEPADLQAIIPGLIGVALLNMTFAFYVAELMGWSVMSLMWG--FA 172
 DB 95 PIFGGRWTVFSTAI-LIIPCVWLGIVAVONNPTFGIFVIALCG-----PAGANFA 146
 QY 173 NEITKIH---EAKRFYALFGIGANISILASGRAIVASKLRAVSSEGVDFWGISLRLM 228

DB 147 LSMGNISFPFPAKAGSAL-GINGGL-----GNLGVSMQVLAFLVIFVPEFA 193
 QY 229 AMTI-----VSGIVMAASYW-----WINKVULDPRTYNEEMKGGKA 268
 DB 194 FLGNGVPOADGWSVMSJANAMWVPLLAITATAMSGMNDI-----A 236
 QY 269 KPRKNNMDSFLYLDNRSPYILTLTLVIA-YGICINLIEVYWSQKLQYPMN-----DY 322
 DB 237 SSRASINDQLPVLQRL-HMLISLTYATFGSPFG-PSAGFAMIAKQFPDVLIRLAF 294
 QY 323 SEFWNGFSFWTGVVSVLIMLVFGNVIRKFGWLTGALVTPVMVL-----TG-- 369
 DB 295 GPFIG-----AIASVSGAISDKFGVAVTLINFLPMAIFGALLFLTPGSG 343
 QY 370 --IVFPALVIFRNDASGLVAMFGTTPMLAVVGAIONISKSTKY--ALPDKEMAYI 425
 DB 344 NFIAFYAVFNGFLFTALGSG--GSTFGMIVIFROI-----TIVVKKGGSDQAH- 393
 QY 426 PLDEQKVKGAADIVVAARFGKSGALIQGL---LVIGS--IGANTPYLAIVLFLIA 481
 DB 394 ---KEAVTEFAALGFTSA-IGAVGFFIPQARMSLMTGSPVGAANKVFLIFIVCVL 449
 QY 482 IWLIV 485
 DB 450 TWLV 453

RESULT 19
 COX1_TRIRU STANDARD; PRT; 528 AA.
 AC 00155;
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 GN COX1.
 OS Trichophyton rubrum.
 OG Mitochondrion.
 OC Bacteria; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Onygenales; Arthrodermataceae; mitosporic Arthrodermataceae;
 OC Trichophyton.
 OC NCBI_TaxID=5551;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IP 1817.89;
 RX MEDLINE=92405249; PubMed=1326416;
 RA de Bievre C., Dujon B.;
 RT "Mitochondrial DNA sequence analysis of the cytochrome oxidase
 RT subunit I and II genes, the ATPase9 gene, the NADH dehydrogenase ND4L
 RT and ND5 gene complex, and the glutamyl1, methionyl and arginyl tRNA
 RT genes from *Trichophyton rubrum*.";
 RL Curr. Genet. 22:229-234(1992).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 CC c + 2 H(2)O.
 CC -1- PATHWAY: Respiratory chain; terminal step.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

FT TRANSMEM 257 277 POTENTIAL.
 FT TRANSMEM 312 332 POTENTIAL.
 FT TRANSMEM 343 363 POTENTIAL.
 FT TRANSMEM 395 415 POTENTIAL.
 FT TRANSMEM 441 461 POTENTIAL.
 FT TRANSMEM 464 484 POTENTIAL.
 SQ SEQUENCE 512 AA; 56122 MW; FID8B1EEB30F1FF CRC64;

Query Match 4.8%; Score 126.5; DB 1; Length 512;
 Best Local Similarity 19.5%; Ered. No. 0.38;
 Matches 109; Conservative 82; Mismatches 169; Indels 201; Gaps 27;

QY 68 FMIVV---PCAIIFMLIYA-KLSNI-LSKQAL-----FYAVGNPFLIFALPFTVYYP 115
 DB 9 FMIVAITAAVINGVISPDLQNVSGAQAIFIDSEFMVLLVLSFVGFCLP-LIFSP 67
 QY 116 LRDVL---HPTPADRLQALPPLGILGVAIIRMTFAFYVLAELMSVMSLMEFWG 170
 DB 68 IGKIKLGPDKPEPF-----GLISWPAFL---FSA-----GMGIGLVFYG 104
 QY 171 FANFI-----TKIHEAKRYALFGIGANI-----S 195
 DB 105 AAEPIISHVAISSPGSETTPOAFDALKYTFEHHGLHMAIYAIYALCIAVFOFRKAPG 164
 QY 196 LLSGRAIIVWASKLRASVSEGV-----PMGISLRLLMAMTIVSG----- 235
 DB 165 LISSTLSPILDKXNGPIGKALDCAVPAIVGVSTSLGATQINGANLFGIPNAFI 224
 QY 236 -----LVMAIYMWINKNVLTDRFYNPBEQKKGAKPKNMKDSFLYLDSP 285
 DB 225 VQVLIIIVVLFILSAM-----SGLGKGIK-----YLSNTN 256
 QY 286 YI-----LLTLTIVANGICI-----NLIEVTWSQKLQYPMNDYSEMGN 328
 DB 257 MVLAGLMLLELVGPVLLINNSTDISQYIQNIIVQSPFLT-----PNDPEKREWINS 311
 QY 329 FSF-----MTGVSVLIMLFVGNVIRKFGMLTGALTVPVWVLLTGIFPALVIFRN 380
 DB 312 WTIIFYMAWMTSMSPFVGIFIAVSRGRTRIF-LIGVLTVPCLITPLMPSIFGVASMDL 369
 QY 381 QASGL-----VAMFGTT--PL-----MLAVVGAIONILSKTKVALPDSTKEMA 423
 DB 370 QOKGAFVNAKLTSTETMFGTLIDHYPLFMVTSILALILIAVFISADS--ATFVLGMQTS 427
 QY 424 YIPLDGQKXKKGKAI--DVVAARFGSGGALLIOGGLVIGSGIGANTPIYALVLLTIIA 481
 DB 428 YGSLNPANSVYLSNGIIOQAAMAVALVYSGIALALONTAIL-----AALPFSIVILLMIAS 482
 QY 482 IW-----LVSATKLNK 492
 DB 483 LYOSLSKERREIKAKELDK 502

RESULT 21
 COXI CYACA STANDARD: PRT: 526 AA.
 AC P48867;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 GN COXI OR COXI.
 OS Cyandidium caldarium.
 OC Mitochondrion.
 OC Bakaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
 OC Cyandidium.
 OX NCBI_Taxid=2771;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RK-1;
 RA Vehmman S.;
 RL Theiss (1995), Justus Liebig University / Frankfurt, Germany.
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY

CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 CC c + 2 H(2)O.
 CC -1- PATHWAY: Respiratory chain; terminal step.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).

CC -----
 CC EMBL, Z48930; CA88773.1; -;
 CC PIR, S62763; S62763.
 CC HSSP, P98002; IARL.
 CC InterPro, IPR000883; COX1.
 CC Pfam, PF00115; COX1; 1.
 CC PRINTS, PR01165; CYCOXIDASE1.
 CC PROSITE, PS00077; COX1; 1.
 CC Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 CC Respiratory chain; Inner membrane.
 CC METAL 65 65 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 CC METAL 244 244 COPPER B (PROBABLE).
 CC METAL 248 248 COPPER B (PROBABLE).
 CC METAL 293 293 COPPER B (PROBABLE).
 CC METAL 294 294 COPPER B (PROBABLE).
 CC METAL 379 379 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).
 CC METAL 381 381 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 CC FT CROSSLINK 244 248 1'-histidyl-3'-tyrosine (By similarity).
 CC SQ SEQUENCE 526 AA; 58354 MW; 5A6CA16B336901B6 CRC64;

Query Match 4.8%; Score 126.5; DB 1; Length 526;
 Best Local Similarity 18.4%; Ered. No. 0.39;
 Matches 89; Conservative 70; Mismatches 160; Indels 165; Gaps 21;

QY 19 IHTHELKKV-----PMEFMFCIT-----FNYVLRDTKTLIYVAGPSGAELPIFK 67
 DB 52 IGHNOLVNVLTAAHGLILIFEMVITPLMGSGFNMV-----PLIIGAPDMAEPRLNNIS 105
 QY 68 FMIVVPCALIFM-----LIYAKLSNLSQALFAVAGTFFLFPALFPTVY 114
 DB 106 FMIMPSLILLLASAFVETGAGTGWTLVPLSSV--QA--HSGGAVDLAIFSLH--IS 157
 QY 115 PLRDVLPTEFADRLQALIPPG-----LIGLVAIL---R 145
 DB 158 GISSILGASNFATIFNIRNPQGNVIRPLFWVSUVTATIIILTPPVLAGATITILLTD 217
 QY 146 NMTFAAFYVLAELMSVMSLMEFW-----GFANEITKHEAKRYALFG- 189
 DB 218 NFN-TSFFDSGGADPVLFQHLFWFGHPEVYILVPAFGIISQVSTFSRKKVFGVGI 276
 QY 190 IGANISLISGRAIIV-----ASTLRASVSEGV-----DPWGISL 224
 DB 277 IYALISIRILG-SWVNAHMFITGMDVTRAYTAAASLILAVPTGIVSVWITMVGSI 335
 QY 225 RLIMANTIVSGLVMAIYMWINKNVLTDRFYNPBEQKKGAKPKNMKDSFLYLD 284
 DB 336 SLKTPMLFAIGFIIIFVGSGLVAVANSGL-----DISLHDYVVAHF 380
 QY 285 PYILLTLVIANVCITNILEVTWSQKLQYPMNDYSEFMGNFSFGVSVLIMLF- 343
 DB 381 HYVLISGAL---FGIFAGF---YYWIKI-----CGKQYSETLIQIHFMTIFGIVNLTFFP 430
 QY 344 -----VGNVIRKFGWMLTGALTVPVWVLLTGIVPALVIFPNQAS 383

```

Db      431 MHFGLAGIPRIIPDPYDAVEEMNIVSTYTG-----AKSIIIGTLIFVYVLAFTN 481
Qy      384 GLVA 387
        ||:|
Db      482 GLIS 485

RESULT 22
PTSA_STRMU STANDARD; PRT; 664 AA.
AC P12655;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PTS system, sucrose-specific IIAB component (IIAB-C-SCR) (Sucrose-
DE permease IIAB component) (Phosphotransferase enzyme II, ABC
DE component) (EC 2.7.1.69) (BII-SCR).
GN SCRA OR SMU.1841.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=89123027; PubMed=2536656;
RA Sato Y., Foy P., Jacobson G.R., Kuramitsu H.K.;
RT "Characterization and sequence analysis of the scra gene encoding
RT phosphoenolpyruvate-dependent sucrose phosphotransferase system.";
RL J. Bacteriol. 171:263-271(1989)..
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
RN [3]
RP SEQUENCE OF 639-664 FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=93329360; PubMed=8336109;
RA Sato Y., Yamamoto Y., Kizaki H., Kuramitsu H.K.;
RT "Isolation, characterization and sequence analysis of the scrc gene
RT encoding fructokinase of Streptococcus mutans.";
RL J. Gen. Microbiol. 139:921-927(1993).
CC -| FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC AND THE TRANSMEMBRANE CHANNEL, THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -| CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC histidine + sugar phosphate.
CC -| SUBCELLULAR LOCATION: Integral membrane protein.
CC -| SIMILARITY: Contains 1 PTS EIIA domain.
CC -| SIMILARITY: Contains 1 PTS EIIIB domain.
CC -| SIMILARITY: Contains 1 PTS EIIC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M22711; AAA26971.1; -.

```

```

DR EMBL; AE015011; AAN59464.1; -.
DR EMBL; D13175; BAA02466.1; -.
DR PIR; B32243; B32243.
DR HSSP; P08637; IGIC.
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001966; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA.1; 1.
DR Pfam; PF00367; PTS_EIIB.1.
DR Pfam; PF02378; PTS_EIIC.1.
DR Pfam; PF002243; PTS_EIIA.1.
DR Pfam; PF001476; PTS_EIIB.1.
DR TIGRFAMs; TIGR00826; EIIB_GIC.1.
DR TIGRFAMs; TIGR00830; PTBA.1.
DR PROSITE; PS00371; PTS_EIIA.1; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS.1.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Phosphorylation; Transmembrane; Complete proteome.
FT DOMAIN 1 40 EIIB DOMAIN.
FT DOMAIN ? ? EIIA DOMAIN.
FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 331 331 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 585 585 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 664 AA; 69988 MW; 809E63E32281A9A1 CRC64;

Query Match 4.8%; Score 126; DB 1; Length 664;
Best Local Similarity 19.9%; Pred. No. 0.53; Indels 200; Gaps 26;
Matches 112; Conservative 83; Mismatches 168;

Qy 16 LMPHTEHLKVLKVLMPFLMFCITFPNYVLBDTKDLIVGARGSGAEAIPIKFWLVPCA 75
Db 82 LTVSTDDDKKIASGKKFNPIMALKLSDI-----FPIIPAL 121

Qy 76 IIFMLYAKLSNLSKQALFYAVGTPPLIFPAEPFVIYPLRDVLPTEPADRIQALPP 135
Db 122 VAGGLVMA-LNNPFLTSEGLF---GTSKLV--QCPPII-----KGSDDMIQ----- 160

Qy 136 GLGLVAILNNWTAATVYVLAELMGSVMLSMFEGFANEITKIHAKRFALPFIGANIS 195
Db 161 -----LMSAPFW-FLPITVGSAAKRFGANQFLGASIG 193

Qy 196 LL--ASGRATWASKRLRASVSEGVDP-----MGI-SLRLLMNTIVSGVLMAVWYIN 246
Db 194 MIMVAPGANIIGLAANAPISKATIGATYGFMNIPBLHTQASVYQVPLVAVVLL- 252

Qy 247 KNVLTDRFYNPBEEMQKKGAKPKMMKDSFLYLDSPYIL--LTLVLVANGICINLI 304
Db 253 -SIL--EKFPF-----KRLPSAVDFPFTPLSLVITNGFLTFIVIG----- 289

Qy 305 EVTMKSQKIQPYMNDYSEFMGNFSFW-----TGVSVVLMLFTGNGVIRKFGMLTGALV 360
Db 290 -----PVMKVSQWLTNGIVMLYDTTG-----FLGKGVF-----GALY 322

Qy 361 TPVWVLLTGI--VEFALVIFRNQASGLVAMF--GTPPLMLAVVGAIONLSKSTKALP 416
Db 323 SP--VWVTGLHQSPPAL-----ETQLISAFOQNTGSGDIFPVASMANVAQGAATPAIY 374

Qy 417 DSTKEMAYIPLDQKQYKGAIDVVAARFGSGGALIQGGL---VIGSGIGAMTPYL 472
Db 375 FLTK-----DKKKKGSSSGVSGVAGLIGTEPALFGVNLKXRPFPFALLIGSASAAA 425

Qy 473 AVILLFIATWVSA-----TKNKFLA 496
Db 426 IAGLLQVAVASLSAGLGLSLKASSIPRYVVCCLISPAIAVAVYGYGKTYAVDVFAA 485

Qy 497 QSALKE-----QEVQEDSAPAS 514
Db 486 EAAVEAIEVQELPEBAASAN 508

RESULT 23
RAFP_PEDPE

```

```

ID RAPE PEDPE STANDARD: PRT: 641 AA.
AC P43466:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Raffinose carrier protein (Raffinose permease).
DE RAPE.
OS Pediococcus pentosaceus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae; Pediococcus.
OK NCBI_TaxID=1255;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=PEP1.0;
RA Leenhouts K.K.J., Bolhuis A.A., Kok J.J., Venema G.G.;
RL Submitted (XXX-1994) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC SODIUM:GALACTOSIDE SYMPORTER FAMILY (SGP).
CC -1- SIMILARITY: Contains 1 PTS EIIA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: Z32771; CAA83664.1; -.
CC EMBL: L32093; AAA25563.1; -.
CC PIR: S44253; S44253.
CC HSP: P20166; IGPR.
CC InterPro: IPR001927; Na/Gal_symport.
CC InterPro: IPR001127; PTS_EIIA.
CC Pfam: PF00358; PTS_EIIA_1; 1.
CC ProDom: PD002243; PTS_EIIA_1.
CC TIGRFAMS: TIGR00792; gdh; 1.
CC TIGRFAMS: TIGR00830; PTBA; 1.
CC PROSITE: PS00371; PTS_EIIA_1; 1.
CC PROSITE: PS00872; NA_GALACTOSIDE_SYMPT; 1.
CC Transmembrane; Sugar transport; Transport; Symport; Phosphorylation.
CC DOMAIN 1 ? SGF DOMAIN ?
CC FT MOD_RES 559 559 PHOSPHORYLATION (BY HPR) (BY SIMILARITY).
CC FT TRANSMEM 25 45 POTENTIAL.
CC FT TRANSMEM 57 77 POTENTIAL.
CC FT TRANSMEM 93 113 POTENTIAL.
CC FT TRANSMEM 120 140 POTENTIAL.
CC FT TRANSMEM 168 188 POTENTIAL.
CC FT TRANSMEM 201 221 POTENTIAL.
CC FT TRANSMEM 253 273 POTENTIAL.
CC FT TRANSMEM 288 308 POTENTIAL.
CC FT TRANSMEM 317 337 POTENTIAL.
CC FT TRANSMEM 342 362 POTENTIAL.
CC FT TRANSMEM 394 414 POTENTIAL.
CC FT TRANSMEM 429 449 POTENTIAL.
CC FT TRANSMEM 485 505 POTENTIAL.
CC FT TRANSMEM 641 AA; 69913 MW; 15BEC569F8C0F61 CXC64;
SQ SEQUENCE
Query Match 4.8%; Score 125.5; DB 1; Length 641;
Best Local Similarity 19.1%; Pred. No. 0.55;
Matches 92; Conservative 83; Mismatches 156; Indels 151; Gaps 24;
OY WLWV---PCALIFMLIVAKSLNIIISKALFFAVGTPPLIFALPPT--VIVPLADV----- 119
DB 93 WVVVGGVYSSIIILLLFTNLGGLYAKNMITYV---VFALIVYTHDIFSPYKOVGFW 147
OY 120 LHPTEPAD-----RLQALIPGLIGLIVAILRNWTFAPAFVIALMGSVWLSIMF 168
DB 148 MLPSLTDSRERKATPARGSTIGGLVGVTL-----VMPAVIF 187
OY 169 WGPANETKIKHEAKRFYVALFGIGANISLAKGRALWASKLRASVSGVDPMWGISLRLM 228

```

| | | | | |
|----|-----|---|----------------------------|-----|
| Db | 188 | --FSKAKSTGDNRCMFIFALICILALISA----- | WGVGIG--- | 222 |
| Qy | 229 | AMTIVSGLVLMASWYWNKNVLTDPFPYNEEMQKCKAKPRKNNKDSPLYIDRSPLYL | | 288 |
| Db | 223 | -----TREVDSDIR----- | KKKQDTVGWMEIFKALAKNDQLL | 254 |
| Qy | 289 | LTLTLVIVAGICINLIEVTWKSQQLQYPPNNNDYSEPMKN----- | FSFWTGVVSVL | 339 |
| Db | 295 | WAAALAYLVYGVGINLT-----GSLEVVY----- | FTYIMKPKPSFSILSIINIFLGIATS | 304 |
| Qy | 340 | IMLFVGWNVIRKFGMLTALVTPVWVLLTGIVFPAIVIPRNOASG----- | VAMEGTTP | 393 |
| Db | 305 | LFPVLSKFRSK-GRFACLV-----FMIGLIALFITA----- | GSNLMVLAAATMGCFPQ | 354 |
| Qy | 394 | LMALVVGAIIONLSKSTKYALF-----DSTKEMAYIPL-DOEQKYKKAIDVVAARFG | | 447 |
| Db | 355 | QMFVLV---LMVITDSVEYGGQLKGRHDEBSLATSVPRLIDRFGAISNGVQGMISG | | 411 |
| Qy | 448 | KSGGALLIQGLLVITGSGIG-AMTPLYAVILFLTIAIMVLSATKLNKFLAQSALKEQEA | | 506 |
| Db | 412 | MTTGATGAS--ITAGQHFKLTFPAPALMLLIAIGPS----- | KQIFLTER--KHAEIV | 463 |
| Qy | 507 | QE 508 | | |
| Db | 464 | AE 465 | | |

Db 122 TFAFAKLMQOGKFIIGALLIS---IIILFFIIVPTVALFISWFDGETAPRQVRI 178
 Qy 80 ----LIYAKSNILSKQALFYAVCTPPLIFPALFPTVYPLRDVHHPFEDRLQAL-- 133
 Db 179 LGQGYIVRVINSLFLSGFVIGIVTFGLAFALYTT-----RLAHRAFAFGKIFSLPI 232
 Qy 134 --PGLGLVAIL-----RWTFAPFV--LAELMGVMSLMFMG 170
 Db 233 VTFPFFVGLGVTLMGRSGVTEFLDTYFGFRDHNMLVGFNGIALAQLAPRPSFMLD 292
 Qy 171 FANEITKIH-----EAKRFYALFGI-----GANIS--LLA 198
 Db 293 GA--LKSHPSEIEASVYTLRANRYQTFFYNILFPLRLPALNSFLIVFIQSLADSPNPLVL 350
 Qy 199 SGRALIVASKLRASVSEGVDPWGISLRLMAMTVISGVLMASVMT--NKAVLDDPRFN 257
 Db 351 GGSFVDVATQIFYIYAGSQLDYASASTGSMILFSLILFIYQVMIGNRSYVT----- 404
 Qy 258 PEEMQKGGKGAQPKNMKDSFLYLDSPYILLTLVVI---AVG-ICINLIEVTWKSQ 312
 Db 405 ----VSGKSYRGVDODLPFGAKY-----TIYMLGFWVFNFALYGSIFVYSGFTVMGVY 456
 Qy 313 KLOYPNMNDYSEFMG--NFSFWGVSVLIMLVGNVIRKFGMLTGALV----- 360
 Db 457 TL--TLNNYAMLFGQGLSDGAMPSLINTMIYAGIAPLTLFGLLIVIVVRKDFQKK 513
 Qy 361 -----TPVAVLTGIVFPLVIFRNQASGLVAMFGTTP 393
 Db 514 TLEFLTMLCFAPVPTVAGSVYLAENDAPMYITGIIIIISMWRDLPIGKRAIAG-- 571
 Qy 394 LMLAVVGAIONILSKTKVALPDSTKEMAYI--PLDOEQKVGK-----KAALDVVAR 445
 Db 572 ----LQGLDKSLDEASLSLKSGSIKTIWTFIVFLKPALSLALSVYFPAAMTVSAIV 625
 Qy 446 FGSAGGALLIOGGLI---VIGSIGAMTPYLAIVLITAI-----WIVSATKLNK 492
 Db 626 FLVYADTRVATAYILNRVEDGEYVALVYGSILIVVMAAILFPDVIAGDTRIRK 680

RESULT 26
 YN86_YEAST STANDARD; PRT; 894 AA.
 ID YN86_YEAST STANDARD; PRT; 894 AA.
 AC P27514;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical 99.5 kDa protein in URK1-SWMI intergenic region.
 GN YNR013C OR N2052.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9288c / FY1679;
 RX MEDLINE=95208356; PubMed=7900425;
 RA Verhaaselt P., Aert R., Voet M., Volckaert G.;
 RT "Twelve open reading frames revealed in the 23.6 kb segment flanking
 the centromere on the Saccharomyces cerevisiae chromosome XIV right
 arm.";
 RT Yeast 10:1355-1361 (1994).
 RL [2]
 RP SEQUENCE OF 1-293 FROM N.A.
 RA Maurer C.T.C., Urbanus J.H.M., Planta R.J.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 784-894 FROM N.A.
 RC STRAIN=ATCC 28383 / FL100;
 RX MEDLINE=90384830; PubMed=2169608;
 RA Kern L.;
 RT "The URK1 gene of Saccharomyces cerevisiae encoding uridine kinase.";
 RL Nucleic Acids Res. 18:5279-5279 (1990).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO THE SLC13A FAMILY OF TRANSPORTERS.
 CC PH087 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X77395; CA54581.1; -
 CC EMBL; Z71628; CA96290.1; -
 CC EMBL; Z71629; CA96292.1; -
 CC EMBL; X53998; CA37947.1; ALT_INIT.
 CC PIR; S45135; S45135.
 CC SCD; S0005296; YNR013C.
 CC GO; GO:0015114; P:phosphate transporter activity; IGI.
 CC GO; GO:0006817; P:phosphate transport; IGI.
 CC InterPro; IPR001898; Na/sul_symp.
 CC InterPro; IPR004331; SPX.
 CC Pfam; PF00939; Na_sulph_symp; 1.
 CC DR Pfam; PF03105; SPX; 1.
 CC KW Hypothetical protein; Transmembrane.
 CC TRANSMEM 430 450
 CC TRANSMEM 474 494 POTENTIAL.
 CC TRANSMEM 511 531 POTENTIAL.
 CC TRANSMEM 557 577 POTENTIAL.
 CC TRANSMEM 602 622 POTENTIAL.
 CC TRANSMEM 642 662 POTENTIAL.
 CC TRANSMEM 682 702 POTENTIAL.
 CC TRANSMEM 706 726 POTENTIAL.
 CC TRANSMEM 738 758 POTENTIAL.
 CC TRANSMEM 777 797 POTENTIAL.
 CC TRANSMEM 799 819 POTENTIAL.
 CC TRANSMEM 824 844 POTENTIAL.
 CC TRANSMEM 874 894 POTENTIAL.
 CC SQ SEQUENCE 894 AA; 99490 MM; 21938585EB0509 CRC64;
 Query Match 4.6%; Score 121.5; DB 1; Length 894;
 Best Local Similarity 18.9%; Pred. No. 1.4;
 Matches 116; Conservative 83; Mismatches 178; Indels 237; Gaps 31;
 Qy 3 KTESKPFGRKRSFLMPPIHTEHKKVLPMLMFCITTNVYVLRDVKOTLVGAFSG-- 59
 Db 365 KTDKKKFSKLSS-----SQLRP-----NANIT-----BSMMSGGAGIHA 400
 Qy 60 --EAAPF-----IKFWLVPCAIIFM----- 79
 Db 401 PSTDSLFRRLHMLPKQMLQFTMGQTSLLKFLUTISCFIALTFNLTPPTQDSLQKNCF 460
 Qy 80 --LIYAKSNILSKQALFY-avgTPPLIFPALFPTVYPLRDVHHPFEDRLQALIPPG 136
 Db 461 ALIYASLWATERIPLFVTSIMIPILI--VFPVINDPITS--QPMSPDSSQ----- 510
 Qy 137 LIGLIVALLRWTPRAFPVLAELMGVMSLMFMGF--ANEITKHEAK--RFVALFGIGA 192
 Db 511 -----FILSTWMSVIM-LLLGGFTLAALSIXKNIAKIVLSTHILASAGT 553
 Qy 193 N-----ISLLASGRAIVWASKLRASV--SEGVDPWGISRL-----MAMTVISG 235
 Db 554 NPHFILLTMFVALFVS---MWSVNAAPVLCSIVP---LRLTLPRNCSTAKALLIG 606
 Qy 236 LVIMASYWMIKNVLTDPREFYNEPMQKGGKAK-----P 270
 Db 607 IALASNIGWSPSPASQNIIFSIGIMDPSPSWAEWMFIALPVCFICMAITWILLITFP 666
 Qy 271 KMMKDSFLYLDSPYILLTLVIVAGICINLIEVTW--KSQLKQYPM----- 319
 Db 667 EPVVKILQHPSRDPLTKQMFVTL--VCIITI-VLMCSNDSIGIFGEMGIISIPIV 722
 Qy 320 -----NDSEFNGNFSFMTGVSVVIMLVFGVGNVIRKFGMLTGALVTP----- 362

Db 723 VFPGNGLLTSDPNNFM-----WT-----IVVLAMGFTLLGKAVSSSGLSTMAQLIAQ 772
 QY 363 -----VMULLTGIVFPALVIFRNQASGLVAMGTTPLM-----LAVVVG 401
 Db 773 VEHBPFIIVLIFGLVILVMAVF--VSHTVAMMIYVLMSEIGSNLPSGHSRLIVIA 829
 QY 402 AIONILSKSTKVALPDSKEMAVYPLDQEQVKGAIDVAAAFKSGGALLIOQGLLVI 461
 Db 830 A---LLCGSA-----MGLPTSGFPNVTALSMIDVGDXY-----LIVGTRIT 868
 QY 462 CGSIGAMTPYLAIV 475
 Db 869 RGVPAISLSYAIV 882

RESULT 27
 COX1 ALUMA STANDARD; PRT; 536 AA.
 AC P80430;
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 GN COX1.
 OS Allomyces macrogynus.
 OC Mitochondrion.
 OC Eukaryota; Fungi; Chytridiomycota; Blastocladiaceae; Blastocladiaceae;
 OC Allomyces.
 OX NCBI_TaxID=28583;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 46923 / BURMA 3-35 (350C);
 RX MEDLINE=96081490; PubMed=7490780;
 RA Paguin B., Forget L., Roewer I., Lang B.F.;
 RT "Molecular phylogeny of Allomyces macrogynus: congruency between
 RT nuclear ribosomal RNA- and mitochondrial protein-based trees.";
 RL J. Mol. Evol. 41:657-665(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 46923 / BURMA 3-35 (350C);
 RX MEDLINE=9626032; PubMed=8636971;
 RA Paguin B., Lang B.F.;
 RT "The mitochondrial DNA of Allomyces macrogynus: the complete genomic
 RT sequence from an ancestral fungus";
 RL J. Mol. Biol. 255:688-701(1996)
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO 1 IS THE
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 CC c + 2 H(2)O.
 CC -1- PATHWAY: Respiratory chain; terminal step.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U41288; AAC49234.1; -.
 DR PIR, 863651, S63651.
 DR HSSP, P98002; IARI.
 DR InterPro, IPR000883; COX1.
 DR Pfam, PF00115; COX1; 1.
 DR PRINTS, PR01165; CYCOXIDASE1.

DR PROSITE; PS00077; COX1; 1.
 KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 KM Respiratory chain; Inner membrane.
 FT METAL 67 67 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT METAL 246 246 COPPER B (PROBABLE).
 FT METAL 250 250 COPPER B (PROBABLE).
 FT METAL 295 295 COPPER B (PROBABLE).
 FT METAL 296 296 COPPER B (PROBABLE).
 FT METAL 381 381 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).
 FT METAL 383 383 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT METAL 246 250 1'-histidyl-3'-tyrosine (By similarity).
 SQ SEQUENCE 536 AA; 59444 MW; 7DE8DF291EAE4091 CRC64;
 Query Match 4.6%; Score 121; DB 1; Length 536;
 Best Local Similarity 20.4%; Pred. No. 0.88; Indels 144; Gaps 24;
 Matches 93; Conservative 65; Mismatches 155;
 QY 22 HELKKVL---PMFLMPCITF-----NYTVLRDTKDTLIVGAPSGAEALPIFIKFWL 70
 Db 57 HQLVNVIITAHAFIMIFFLVMPMLGSGFNYPV-----PIMIGAPDMARPLNNISFWL 110
 QY 71 VVPCALIFM-----LIYAKLSNLSKQALFYAVGTFPLIFPL----- 108
 Db 111 LPPLILLVGSAFVGEQAGTGTVPYPLSSI-----GFHSGGSVDLAIFSLHLAIGISML 165
 QY 109 -----FPTVIYPLRD---VLHPTB---PADRLQAL---PPGLLGVAII---RNTFFAA 151
 Db 166 GSINFIITLLMRAFGMTWKPLPLFVWSILITLILSLPVLGALITMLLTDRLNL--TT 224
 QY 152 FYVLAEIAGSVMLSLMFW-----GFANEITKIHAKRFVALFGI---GAN 193
 Db 225 FYDPAGGDPVLVQHLMFPGHPEVYIIIRGGIISQVSTSRKRIFFYLGAVVAMAS 284
 QY 194 ISLASGRAVYASKLRASVSEGVDPWGISRLMANTYISGLVMASTYMT---NKVY 249
 Db 285 IGLIG---FIVWGHM--YTVGLDVTDFAYEPTATMTIAVPTGKIFS---WLATLYGNI 337
 QY 250 LTPDRFYNPEMKG-----KKGAKPKNMDSFLYDRSPYILL--TLV 294
 Db 338 L---YTPAYFALGFLFTTGCVGWMLANSLDALDITTVYVAAHFYVLSWGAIFA 393
 QY 295 IAVGICINLIEVTKSQLKQYPMNDYSEFGNPFMTGVSYL---MLFVG----- 345
 Db 394 LPAGFYWIKIGTK-----QYNFEGQVHWMTFIGNVTVFFPMHFLGLNGMP 442
 QY 346 -----GNVIRKFGMLGALVTPVMVLTGI 370
 Db 443 RRPDPYDAFTQNNVVISFSGISISIVST--IVELYGL 477

RESULT 28
 YDHP_ECOLI STANDARD; PRT; 389 AA.
 AC P77389;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical transport protein ydhp.
 GN YDHP OR B1657.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatcher F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Coliado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).

[2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kashi H., Kashiimoto K., Kimura S., Kitakawa M.,
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
 RA Mochimaru K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
 RA Oshima T., Saito N., Sempel G., Seki Y., Sivasubraman S.,
 RA Tagami H., Takeda J., Takekoto K., Takeuchi Y., Wada C.,
 RA Yamamoto Y., Horiuchi T.;
 RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (potential).
 CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
 CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE000261; AAC74729.1; -;
 DR EMBL; D90809; BAA15423.1; -;
 DR EMBL; D90810; BAA15432.1; -;
 DR PIR; C64923; C64923.
 DR EcGene; EG13950; ydhp.
 DR InterPro; IPR007114; MFS.
 DR InterPro; IPR005828; Sub_Transporter.
 DR Pfam; PF00083; sugat_tr; 1.
 KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
 KM Complete proteome.
 FT TRANSMEM 7 27 POTENTIAL.
 FT TRANSMEM 44 64 POTENTIAL.
 FT TRANSMEM 71 91 POTENTIAL.
 FT TRANSMEM 101 121 POTENTIAL.
 FT TRANSMEM 131 151 POTENTIAL.
 FT TRANSMEM 160 180 POTENTIAL.
 FT TRANSMEM 204 224 POTENTIAL.
 FT TRANSMEM 237 257 POTENTIAL.
 FT TRANSMEM 267 287 POTENTIAL.
 FT TRANSMEM 291 311 POTENTIAL.
 FT TRANSMEM 331 351 POTENTIAL.
 FT TRANSMEM 357 377 POTENTIAL.
 SQ SEQUENCE 389 AA; 40064 MW; 90A6A22CDB565B3 CRC64;
 Query Match 4.6%; Score 120; DB 1; Length 389;
 Best Local Similarity 22.8%; Pred. No. 0.75;
 Matches 109; Conservative 62; Mismatches 167; Indels 140; Gaps 25;
 QY 40 NYTVYRTKOTLVGARGSGAE-----AIPFIKFWLVVPCATFMILYATLSNLSKQA 93
 DB 4 NYPL-----ALATGARGIGITTEPSPWGLFVIRAGVDVSIIPAAGMI-----SAYA 50
 QY 94 LFVAVTPPLIFF-----ALFPYIYPLRDVLHPTEPADRLQAILPPGL-LGVAVI 143
 DB 51 VGWVVGAFWMTLLSHRARRSALFLMAIFLGNV-----LSAIPDYMTLMSRI 101
 QY 144 LRNWTFAFYVLAELWGSVMSLWFMWGFANEITYIHEAKRFYALFGIGANISLLASGRAI 203
 DB 102 LITSINHGAFFGL-----GSVVA-----ASVVPKHQASAVATMF-MGLTLANIGVPA 149
 QY 204 VMASKLASVSEGVDPNGISIRLIMAMTIYSGVLMSYWMYINCVLTDPRFYPEEKOK 263
 DB 150 TWLGET-----IGWRMSFLNAGLGVISWVSLF-----SLPK 182
 QY 264 GKXGAKPRMKNKDSFLYLDSPYLLLTLLVIAVIGICINILEVWKSQKLQYENMNDYS 323

DB 183 GGAGARPE--VKKEVLAMRQVLSALLTTLVLAGAMFTL--YTYIS-----PVLOSIT 232
 QY 324 EPMGNSEFWTCVSVLMLFVG---GNVT-----RKFGULTALVTPVWVLLTGIVFF 373
 DB 233 HATPVF-----VTAMLVIGVFSIGVYLGKGLADRSVNGTLKGFLL-LMVIMLAIPL 286
 QY 374 ALVIFRQASGLVAM--FGTT-----PLMAVY-VGAIQNLISKSTKYMALPDSKTWAY 424
 DB 287 A---RUEFGALISVWVGAAATPVVPLQWRVRVASAPGLSSVNGIGFNGLNL-- 340
 QY 425 IPLDQEKVYKKAIDVVAARFGKSGALLIOOGL-LVTCSSIGAMTPYALVILFIIT 481
 DB 341 -----GAA-----GGAVISAGLGYSFVPMGAIVAGLALLVFM 377
 RESULT 29
 YDHP EC057 STANDARD; PRT; 389 AA.
 AC Q8X625;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical transport protein ydhp.
 GN YDHP OR Z2679 OR EC82366.
 OS *Escherichia coli* O157:H7.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=83334;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfale G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potlamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";
 RL Nature 409:529-533(2001).
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishi K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (potential).
 CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
 CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE005389; AAG56646.1; -;
 DR EMBL; AP002558; BAB35789.1; -;
 DR PIR; B85773; B85773.
 DR PIR; F90924; F90924.
 DR InterPro; IPR007114; MFS.
 DR InterPro; IPR005828; Sub_Transporter.
 DR Pfam; PF00083; sugat_tr; 1.
 KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
 KM Complete proteome.

FT TRANSMEM 7 27 POTENTIAL.
 FT TRANSMEM 44 64 POTENTIAL.
 FT TRANSMEM 71 91 POTENTIAL.
 FT TRANSMEM 101 121 POTENTIAL.
 FT TRANSMEM 131 151 POTENTIAL.
 FT TRANSMEM 160 180 POTENTIAL.
 FT TRANSMEM 204 224 POTENTIAL.
 FT TRANSMEM 237 257 POTENTIAL.
 FT TRANSMEM 267 287 POTENTIAL.
 FT TRANSMEM 291 311 POTENTIAL.
 FT TRANSMEM 331 351 POTENTIAL.
 FT TRANSMEM 357 377 POTENTIAL.
 SQ SEQUENCE 389 AA; 40035 MM; B538643270688816 CRC64;

Query Match 4.5%; Score 119; DB 1; Length 389;
 Best Local Similarity 22.8%; Pred. No. 0.86;
 Matches 109; Conservative 62; Mismatches 167; Indels 140; Gaps 25;

QY 40 NYTVLRDRTKTLTGAPSSGAE-----AIPPIKFWLVPCALIFMLYAKLSNLSKQA 93
 DB 4 NYPL-----ALATGAFGIGTTERSPMGLPVIARGVDSIPAGMLI-----SAYA 50
 QY 94 LFVAVGTFPLIFF-----ALFPIYVPLRDVLFTEFADRLQALIPGL-LGLVAI 143
 DB 51 VGWVAGAPLMTLLSHRARSALIFLMAIFLGNV-----LSAIPDYMTLMSRI 101
 QY 144 LRNTFAAFVYLAELMGSVMSLMEFWGANEITIKHEAKRPYALFGIGANISLASGRAI 203
 DB 102 LTSINHGAFFGL-----GSVA-----ASVVPKHKQASAVATWF-MGLTLANIGVPA 149
 QY 204 VMAKRLASVSEGVDPWGISRLMLAMTIVSGVIVMAVSVYMNKVLDPFPVPEEMOK 263
 DB 150 TWLEST-----IGWMSRLATAGGVLSWSLFF-----SLPK 182
 QY 264 GKGAAPKPMNKDFLYDRSPYILLTLVAVGICINLIEVTKSQKLQYPMNDYS 323
 DB 183 GGAARPE--VKKELAVMRPOVLSALTLYLAGAMPTL--YTIIS-----PYLSIT 232
 QY 324 EFMGNFSWTVGVSVLMLFNG--GNVI-----RKFGLTGALVTPVWVLTGIYEF 373
 DB 233 HATPEV-----VTAMLVIGVFGISGNVIGKADRSVNGTLKGFLL-LWVIMLAIPL 286
 QY 374 ALVIFRNASGLVM--FGTT-----PLMLAVV-VGAIONILSSTYVALFSTKEMAY 424
 DB 287 A-----RNFEGALISAVGAATPAVVPLOKRWKVASDAQSLSSVNIKFNIGNAL-- 340
 QY 425 IPLDOEKVKGAIDVVAARFGKSGGALLIOGL-LVIGSIGAMTPYLAIVLLEPIA 481
 DB 341 -----GAAA-----GGAVISAGLGYSVPVWGAIVAGLALLLVFMSA 377

RESULT 30
 NARU SALTY STANDARD; PRT; 462 AA.
 AC P37593;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nitrite extrusion protein 2 (Nitrite facilitator 2).
 OS NARU OR STM1576.
 GN Salmomella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmomella.
 OC NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RA MEDLINE=21534948; PubMed=11677609;
 RA McEllelland W., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;

RT "Complete genome sequence of Salmomella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 RN [2]
 RP SEQUENCE OF 1-130 FROM N.A.
 RC STRAIN=SL1303;
 RX MEDLINE=95011654; PubMed=7926834;
 RA Hongo E., Morimoto M., Mita K., Machida I., Hama-Inaba H., Tsuji H.,
 RA Ichimura S., Noda Y.;
 RT "The methyl viologen-resistance-encoding gene smva of Salmomella
 typhimurium";
 RL Gene 148:173-174(1994).
 CC -1- FUNCTION: INVOLVED IN EXCRETION OF NITRITE PRODUCED BY THE
 CC DISIMILATORY REDUCTION OF NITRATE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (probable).
 CC -1- MISCELLANEOUS: SINCE THE TRANSPORT IS CONTROLLED BY OXYGEN, FIVE
 CC CYS RESIDUES IN NARU MAY PLAY AN IMPORTANT ROLE IN SENSING REDOX
 CC POTENTIAL TO REGULATE THE FACILITATOR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NARU/NASA FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL; AE008769; AAL20494.1; -.
 DR EMBL; D26057; BAA05054.1; -.
 DR Stycogen; SG10245; narU.
 DR InterPro; IPR004737; NO3 transporter.
 DR TIGR; TIGR00886; ZAO108; 1.
 KW Nitrate assimilation; Transport; Transmembrane; Inner membrane;
 KW Multigene family; Complete proteome.
 FT TRANSMEM 36 56 POTENTIAL.
 FT TRANSMEM 72 92 POTENTIAL.
 FT TRANSMEM 102 122 POTENTIAL.
 FT TRANSMEM 125 145 POTENTIAL.
 FT TRANSMEM 181 201 POTENTIAL.
 FT TRANSMEM 207 227 POTENTIAL.
 FT TRANSMEM 259 279 POTENTIAL.
 FT TRANSMEM 288 308 POTENTIAL.
 FT TRANSMEM 318 338 POTENTIAL.
 FT TRANSMEM 342 362 POTENTIAL.
 FT TRANSMEM 402 422 POTENTIAL.
 FT TRANSMEM 433 453 POTENTIAL.
 FT TRANSMEM 461 481 POTENTIAL.
 FT CONFLICT 61 N -> D (IN REF. 2).
 FT CONFLICT 126 TPFGV -> FLISA (IN REF. 2).
 SQ SEQUENCE 462 AA; 49751 MM; 2CDCC677B9480B53 CRC64;

Query Match 4.5%; Score 118; DB 1; Length 462;
 Best Local Similarity 20.7%; Pred. No. 1.2;
 Matches 101; Conservative 66; Mismatches 179; Indels 142; Gaps 23;

QY 69 WLVVPCALIFMLYAKLSNT-----LSKQALFYAVGTPPLIFALPPTIYVPLRDVLPHEF 125
 DB 37 WISVSCILLAFCCWMLBSAAVAVNINKIGFNFTDQFL-----TAL 78
 QY 126 ADRLOALIPGLGLVAAIL--RNWTFPAFVYL-----AELMGSVMSLMEFW 169
 DB 79 PLSGAILRPVPSVWVLPFGKRKTIVSYVILLIPCAMGRVAVGNPATPFVFLIALLC 138
 QY 170 GFANEITIKHEAKRPYALFGIGANISLASGRAIWMASKLR---ASVSEGVDPWGISLR 225
 DB 139 GFA-----GANFA--SSWGNISFFPKARQSGALGINGLGNLGVSVW 179
 QY 226 LLMAMTV-----SGVLMASYW-WIN-KVNLTPDRFPNPEEMOKK 265
 DB 180 QILPVLVIFPLIFPLGVOGVPDGSILALTNAWTVPLAVATLAAMGAMNDIGSSK 239
 QY 266 KGAKPKNMKDSFLYDRSPYILLTLVIA-VGICINLIEVTKSQKLQYPMNDY 320


```

Db 240 ASVA-----SOLPVLRLHLMLSLLYLATFGSFFIG-PSAGFAMLAKTGPDPVNIQL 291
Qy 321 -DYSEEMGNFSPFWGVSVLMLFVGVNVIKFGMLTALTPV-MVLTGTGVFALVIF 378
Db 292 AFGPFPIG-----ALARSAGVISPDEFGVRYLLINFLFMAFLFLLFTL--- 337
Qy 379 RNQASGLVAMF-----GTTPLMAVVGAIQNLISKSTKY--ALPDSYKE 421
Db 338 PGSGAGSFAFVIVFMGLFPLAGLGSGSTFQMLAIFRQI-----TLNVKLRGSGDE 390
Qy 422 MAYPLDGEQYKGAIDVVAARFGKSGALIQGL---LVTCGS-IGAMTPYLAVILL 477
Db 391 QA-----QREAVTDPAALGFISAIGAVGGEFIPKAFGTSIALTSSPVGAMKIFLLFYLA 445
Qy 478 FFIAMTV 485
Db 446 CVLLTWLV 453

RESULT 31
PUTP_HAEIN STANDARD; PRT; 504 AA.
ID PUTP_HAEIN STANDARD; PRT; 504 AA.
AC P45174;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sodium/proline symporter (proline permease).
GN PUTP OR H11352.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN RN
RF SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Lin L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -1- FUNCTION: CATALYZES THE SODIUM-DEPENDENT UPTAKE OF EXTRACELLULAR
CC L-PROLINE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.embnet.ch/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL, U32814; AAC22999.1; -.
CC PIR, E64118; E64118.
CC TIGR, H11352; -.
CC InterPro, IPR001734; Na/solut_symport.
CC Pfam, PF00474; SSF, 1.
CC TIGRFAMs, TIGR00813; sss, 1.
CC PROSITE, PS00456; NA_SOLUT_SYM_1; 1.
CC PROSITE, PS00457; NA_SOLUT_SYM_2; 1.
CC PROSITE, PS50283; NA_SOLUT_SYM_3; 1.
KW Transport; Amino-acid transport; Transmembrane; Inner membrane;

```

```

KW Sodium transport; Symport; Complete proteome.
FT TRANSMEM 8 71 POTENTIAL.
FT TRANSMEM 51 28 POTENTIAL.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 127 147 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
FT TRANSMEM 189 209 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 374 394 POTENTIAL.
FT TRANSMEM 405 425 POTENTIAL.
FT TRANSMEM 434 454 POTENTIAL.
FT TRANSMEM 461 481 POTENTIAL.
SQ SEQUENCE 504 AA; 54898 MW; 61ED0A52A142A7F0 CRC64;

Query Match 4.5%; Score 117.5; DB 1; Length 504;
Best Local Similarity 19.1%; Pred. No. 1.4;
Matches 82; Conservative 77; Mismatches 117; Indels 153; Gaps 22;

Qy 148 TFAAFYLAELMGSVMLSMFWGRANETKHEKRFYALG--IGANISILASGRALVW 205
Db 10 TETTY-----IFGMLLIGVALVYTNMLSD-----YILGRRIGSFVTMSAG----- 52
Qy 206 ASKLRAVSSEGVDPWGISRLMLAMTYVSGVLNASYWM-----INKRVLTDP 253
Db 53 ASDSGMLMLGL-EGAVLSGLVEGWIALIGLTIGAFYMLLVAGRLRYVTELNNAITLP 111
Qy 254 RFPNPEMOKKKAKKAPKMNKDSFLYLDSPYILLTLVAVAGICI-----NLIEVT 307
Db 112 EYFH-----NRFSSHKL-----LKLVSATILVFLTYICAGGVAVGAKLFQNIPEVE 159
Qy 308 WKSQKIQYPMNDYSEPMGNFSPWGVSVLMLFVG-----NVIRFGMLTGALV 360
Db 160 YSTAL-----WGAATAIAYVFIQGFPLAVSWTDTIOATLMIFALIL 200
Qy 361 TPVAVLLT--GIVEFALVIFRQAS--GLVAMF-GTTP-LMLAVVGAI-----QNTLS 408
Db 201 TPVAVLLSFDTPAQSALVBEAANAUNKDFDLTSTTPLGLSLAMGLGFQGPPIILA 260
Qy 409 KSTRYALFDSYTKENAYVPLDGEQYKGAIDVVAARFGKSGALIQGLLVIC--GSIG 466
Db 261 ---RFMAADSVKSL-----IKAR-----RISGMWVLCIAGAG 291
Qy 467 -----AMTPYLAVILFIIAIVLSATYKLU 493
Db 292 IGLFAIPYFANPAIAGTVNREPQVETELAKLFNFWIAGILISAI-LAAVWSTLSAQL 350
Qy 494 FLAQSALKE 502
Db 351 LISSSITE 359

RESULT 32
COXI_MARPO STANDARD; PRT; 522 AA.
ID COXI_MARPO STANDARD; PRT; 522 AA.
AC P26856;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
GN COXI OR COXI.
OS Marchantia polymorpha (Liverwort).
OC Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiopsida; Marchantiales; Marchantiaceae; Marchantiales;
OC Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN RN
RF SEQUENCE FROM N.A.
RX MEDLINE=92114051; PubMed=1731062;
RA Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N.,
RA Akashi K., Kanegae T., Ogura Y., Kohchi T., Ohyama K.;

```

"Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome";

RT J. Mol. Biol. 223:1-7(1992).

CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B.

CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome c + 2 H(2)O.

CC -1- PATHWAY: Respiratory chain; terminal step.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.

CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL, M68829; AAC09451.1; -.

DR PIR, S25956; S25956.

DR HSSP, P00396; 20CC.

DR InterPro, IPR000883; COX1.

DR Pfam, PF00115; COX1; 1.

DR PRINTS, P01165; CYCOXIDASRI.

DR PROSITE, PS00077; COX1; 1.

KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane; Respiratory chain; Inner membrane.

FT METAL 65 65 IRON (HEME A AXIAL LIGAND) (PROBABLE).

FT METAL 244 244 COPPER B (PROBABLE).

FT METAL 248 248 COPPER B (PROBABLE).

FT METAL 293 293 COPPER B (PROBABLE).

FT METAL 294 294 COPPER B (PROBABLE).

FT METAL 379 379 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).

FT METAL 381 381 IRON (HEME A AXIAL LIGAND) (PROBABLE).

FT CROSSLINK 244 248 1'-histidyl-3'-tyrosine (By similarity).

SO SEQUENCE 522 AA; 57551 MW; 4888E16F03A11D01 CRC64;

Query Match 4.4%; Score 116.5; DB 1; Length 522;
Best Local Similarity 19.1%; Pred. No. 1.7;
Matches 90; Conservative 60; Mismatches 157; Indels 165; Gaps 21;

22 HELKKVL---PMFMFCITF-----NTYLRDITDIIVAGPGGAAIPRIKFWL 70
55 HQLNVNLTAHAPLTFPMVPMAMIGFGNMFV-----PLIGSPDMAFPLNNISFWL 108
71 VVPCALIFM-----LTYAKLSNLSKQALFYAVGTPLIFPAL----- 108
109 LPSLLULLSSALVVEGCGSMYTPPLSGITS-----HSGGSVDLAFSLHSVSSIL 163
109 ---PFTVYLPURD---VLH-----PTEPADRLQAILPGLGLVAAIL---RNWT 148
164 GSINFITTFIMMRABGLTMHRLPLFVMSVLTAF---LLLSLPLVLAIGAITMLLTDREFN 220
149 FAAYVVALMELGWSLIMFGFANEIRKI-----HEAKRF-----YA 186
221 -TTFPDPRGGDPLLYOHLFMFPGPEVYLLLPFGGIIISHIVTFSSKPVGYGMYYA 279
187 LFGIGA-----NISLASGRAIVWASKLNASYSEG-----DPWGISLRLLM 228
280 MISIGVLGIYMAHMFVGLDVDTRAVFTAAITMIAIPGKIFSWIATMMGSGIQYKT 339
229 AMTVSGVLVMASTWINKNVLTDRFYNPEBMQKGGKAKPKNMKDSFLYLDSPYIL 288
340 PMLFAVGVIFLFTVGGTLGIVLAN-----SGVDIALHDTYVVAHFHYVL 384

QY 289 LL-TLVAVGICINLIEVTWKSQKLQYPMNDYSEFMGNFSWTVGSVLMLF----- 343
DB 385 SMGAVPALFAGFYWICKIT-----GLQYF-----ETLGGIHMTFFGVNLTFPPWHF 433
QY 344 -----VGVNIRKFGWLGTALVTPVWVLTGLVFALVTF 378
DB 434 LGLAGMPRIPIPDYDAYAGWNAFSGF-----SVYSVGLFCEPVVVF 476

RESULT 33

NRFE_HA6IN STANDARD; PRT; 635 AA.
ID NRFE_HA6IN
AC P44944;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome c-type biogenesis protein nITE.
GN NRFE OR H10936.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
NX NCBI_TaxID=727;
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fieschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bule C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom B., Cotton M.D., Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RT Science 269:496-512(1995).
RL -1- FUNCTION: REQUIRED FOR THE BIOGENESIS OF C-TYPE CYTOCHROMES.
CC POSSIBLE SUBUNIT OF A HEME LYASE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (potential).
CC -1- SIMILARITY: BELONGS TO THE CCMF/CYCK/CCL1/NRFE/CCSA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL, U32775; AAC22594.1; -.
DR PIR, A64162; A64162.
DR TIGR, H10936; -.
DR InterPro, IPR003567; Cyt c biog.
DR InterPro, IPR003568; CytC_biog_CcmF.
DR Pfam, PF01578; CytC_asm.1.
DR PRINTS, P01410; COBIOGENESIS.
DR TIGRFAMs, TIGR00353; nrfe; 1.
KW Cytochrome c-type biogenesis; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 6 26
FT TRANSMEM 38 58 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 167 187 POTENTIAL.
FT TRANSMEM 209 229 POTENTIAL.
FT TRANSMEM 239 259 POTENTIAL.
FT TRANSMEM 273 293 POTENTIAL.
FT TRANSMEM 311 331 POTENTIAL.
FT TRANSMEM 351 371 POTENTIAL.

FT TRANSMEM 390 410 POTENTIAL.
 FT TRANSMEM 421 441 POTENTIAL.
 FT TRANSMEM 452 472 POTENTIAL.
 FT TRANSMEM 483 503 POTENTIAL.
 FT TRANSMEM 570 590 POTENTIAL.
 FT TRANSMEM 607 627 POTENTIAL.
 SQ SEQUENCE 635 AA; 71273 MM; B979E54F1ED534FC CRC64;

Query Match 4.4%; Score 116; DB 1; Length 635;
 Best Local Similarity 19.9%; Pred. No. 2.1;
 Matches 103; Conservative 46; Mismatches 142; Indels 226; Gaps 21;

QY 78 FMILYAKLSNLSKQALFYAVGTPELFPALFPYIVYLDVLPTEPADLQALPGL 137
 DB 8 FLILLATRS-----AFPLAL-VPQGLFKKPKPLI-----NAAPLSITFLATLTLISGL 56
 QY 138 LG-----LVAILRMTFAFYVLAELWGSVMSLMFWGFA----- 172
 DB 57 LAYSFAVDFTLEVVAHNSGQLPFFKVAATWGHGHSMLFMLSLSMLAFAFENRK 116
 QY 173 NEIRKIHAKRFYALFGANISL-----ASGR----- 201
 DB 117 NDRFSAOSLSLGLICGFPAVFILFYSNPGRIFPAPAEGRDLPMLQDVGLIHPPL 176
 QY 202 -----AIWASKLRA-----SVSEGVDPWGISLRILMAMTIYSGVLMASY----- 242
 DB 177 YVGVGVAVNAPMALSALITYQSARQIARSKRGVLSWLPITIGIVG-AMWAIYELGW 235
 QY 243 -----NMINKNVLTDPFPYNEEMOKGKGAKPROMKDSFLYLDSPYILLTLVIAVIGIC 300
 DB 236 GGMV-----FMDEVE----- 245
 QY 301 INLEIVWKSQLKQYPMNDYSEPMGNFSPWGVSVLIMLF-VGVAVIRKFGMLGAL 359
 DB 246 -NASLMPWLLGLALHLSLM--ATERKQGVFSYWTLLFSLAPAFSVLGFIVRSGLTSVH 302
 QY 360 V-----TPVWVLTGIVFPAVIFRNOAGLVAMFGTPLMLAVVGAIQNLSKSTKVA 414
 DB 303 AFALDNRGVYLL-LIFFVLTAL--AFGLPALRA----- 333
 QY 415 LFDSTKEMAYIPLDOEKVKKAIDVVAARF-GKSGGALLIQGGLVLC----- 462
 DB 334 -----GSSSESANVKFQFISGSGILLNLTLTATVSTFLGTFYP 373
 QY 463 -----GSTGAMTPYL-----AVILLFIITAI 482
 DB 374 MLFQAMNWSISVSGSPYENSIFPPLITAILIMIVYL 410

RESULT 34
 YGED ECOLI STANDARD; PRT; 397 AA.
 ID YGED ECOLI
 AC P39196; Q46934;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein yged.
 GN YGED OR B2835.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=9426617; PubMed=9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.F., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RL "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474 (1997).
 RN [2]

RP SEQUENCE OF 1-80 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=94132066; PubMed=8300626;
 RA Jackowski S., Jackson P.D., Rock C.O.;
 RT "Sequence and function of the aas gene in Escherichia coli.";
 RL J. Biol. Chem. 269:2921-2928 (1994).
 RN [3]
 RP IDENTIFICATION.
 RA Rudd K.E.;
 RL Unpublished observations (NOV-1994).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi-sib.ch).

DR EMBL; U29581; AB40482.1; -;
 DR EMBL; AB000367; AAC75874.1; -;
 DR EMBL; L14681; -; NOT ANNOTATED_CDS.
 DR PIR; D65066; D65066.
 DR Ecogene; EG12455; Yged.
 KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 19 39 POTENTIAL.
 FT TRANSMEM 53 73 POTENTIAL.
 FT TRANSMEM 96 116 POTENTIAL.
 FT TRANSMEM 145 165 POTENTIAL.
 FT TRANSMEM 167 187 POTENTIAL.
 FT TRANSMEM 227 247 POTENTIAL.
 FT TRANSMEM 257 277 POTENTIAL.
 FT TRANSMEM 281 301 POTENTIAL.
 FT TRANSMEM 305 325 POTENTIAL.
 FT TRANSMEM 344 364 POTENTIAL.
 FT TRANSMEM 367 387 POTENTIAL.
 FT CONFLICT 75 80 SPKGR -> TSPKAG (IN REF. 2).
 SQ SEQUENCE 397 AA; 41655 MM; 0BB3B0DEA7FF534C CRC64;

Query Match 4.4%; Score 115.5; DB 1; Length 397;
 Best Local Similarity 20.7%; Pred. No. 1.5;
 Matches 86; Conservative 63; Mismatches 143; Indels 123; Gaps 21;

QY 137 LGLVAILRMTFAFYVLAELWGSVMSLMFGFANEIRKIHAKRFYALFG-IGANI 194
 DB 34 LFATLALIK-----AQFY-----PEWSQPILOWFFVG-----AYILFAFPVGVGA 73
 QY 195 SLASGRAIYVASKLR-----ASVSEGVDPWGISLRILMAMTIYSGVLMASYMINKNV 250
 DB 74 DSPAKGVMMFANGLKLGASISICFGINP-----LGITLVG--VGAAVSPAKYGIL 124
 QY 251 TDPFRYNEEMOKGKGAKPROMKDSFL-----YLDSPYILLTLTLVIAVIGI 299
 DB 125 -----GELTTSKLVKANGMEASTIAIILGSAVGVLADWHLVLAACALAVGG 176
 QY 300 CI-----NLIEVT-----WKSQIKIQYPMNDYSEPMGNFSE 331
 DB 177 AVVANIYIPKLAARPGQSNNLIMNTRSPFNACTSLWR-----NGETRRFSLVGTSLF 228
 QY 332 WTVGVSVLIMLFVGVGNVIRKFGMLTGAL-----VTPVWVLTGIVFPAVIFRNOAGSLV 386
 DB 229 WGAGVTIRFLIVL-----WVPVALGITDNATP--TYLMAVAIGIVGACAAAKLV 277
 QY 387 AMFGTPEMLA-VVGAIONILSKSTR-----YALFSTKEMA--YIPLDOEKVKKAA 438
 DB 278 TLEIVSCWMPAGILIGVAVLIFSQHELTPAVYALMLIGWGGFFVPLNALIGERGKS 337
 QY 439 ID-----VVAARFGSGGALLQGGILVCGSIG-AMTP-----YLAVILLFIITAIL 484
 DB 338 VGANALAVONTGNSAMLMILGIYSLAVMIGIPVIVIGIGFGLFALAITALMI 392

RESULT 35
COXI_Apili STANDARD; PRT; 521 AA.
AC P20374;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
GN COI.
OS Apis mellifera ligustica (Common honeybee).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Hymenoptera; Apoecrita; Aculeata; Apoidea;
OC Apidae; Apis.
OX NCBI_TaxID=7469;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thorax;
RX MEDLINE=90136028; PubMed=2559293;
RA Crozier R.H., Crozier Y.C., Mackinlay A.G.;
RT "The CO-I and CO-II region of honeybee mitochondrial DNA: evidence
for variation in insect mitochondrial evolutionary rates.";
RL Mol. Biol. Evol. 6:399-411(1989).
[2]
RN SEQUENCE FROM N.A.
RP TISSUE=Thorax;
RX MEDLINE=93114603; PubMed=8417993;
RA Crozier R.H., Crozier Y.C.;
RT "The mitochondrial genome of the honeybee Apis mellifera: complete
sequence and genome organization.";
RL Genetics 133:97-117(1993).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
AND COPPER B.
CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
c + 2 H(2)O.
CC -1- PATHWAY: Respiratory chain; terminal step.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M23409; AAA18476.1; -;
DR EMBL; L06178; AAB96799.1; -;
DR PIR; A32431; A32431.
DR HSSP; P18401; 1PRT.
DR InterPro; IPR006883; COX1.
DR Pfam; PF00115; COX1.1.
DR PRINTS; PS00165; CYCOXIDASE1.
DR PROSITE; PS00077; COX1.1.
KM Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW Respiratory chain; Inner membrane.
FT METAL 59
FT METAL 238
FT METAL 242
FT METAL 288
FT METAL 289
FT METAL 374
FT METAL 376
FT CROSSLINK 238
SEQUENCE 521 AA; 59293 MW; 2149417AC981CE64 CRC64;

Query Match 4.4%; Score 115; DB 1; Length 521;
Best Local Similarity 20.1%; Pred. No. 2;
Matches 86; Conservative 68; Mismatches 136; Indels 138; Gaps 22;
QY 31 FLMEFCITFNXYT--LRDRTKTLIVGAPGSGAEIPIKFWLVVPCAIIFMLIVAKLSNI 88
DB 61 FLMTFFWMPPLIGFGNWLIPLMLGSPDMAFPKNNISFWLPPS--LEWLT--LSN- 114
QY 89 LSKQALFY-ANGTFELFLFPAFFVIYPLRDVLPHTFADRLQAILPGLGLVAILRNW 147
DB 115 ----LFYSPSGTGVY-----PP-----LSAYVYHS 137
QY 148 TFAAFVYLAEIWSVLMFLMEFGFANEITKHEAKRYALFGIGANISLASGRAIYAS 207
DB 138 SPSSVDFALFSLHNSGISIM--GSLNLMVITIMMKNSMNYD--QSLF----- 182
QY 208 KLRASVSEGVDPWGI---SLRLMAMTIVSGLVIMASYMWINKVLTDPFRYNEPMQKG 264
DB 183 -----PMSVFYTAIILWSLPVLGAILTMLF---DRNFNTS--FFDP----- 220
QY 265 KKGAKPKMKNKDSFLYLDKSPYILLTLVIAVGCINLIEVTMKSQLKQYPMNDYSE 324
DB 221 MGGSDPLLYOHLFEWFGHPEVYIILP---GRELISHIV-----MNSGKKE 264
QY 325 FMGNFSFWTGVSVYLIM-----LFGVGNVIRKFGMLGALVTPYVMTLTGI--VEPAL 375
DB 265 IFGNLSMIVYMLGIFGFIYVAHMHMTVGLDVDTAVFIYSA--TMIAVPTGIKVSWM 322
QY 376 VIFENQ-----ASGLVAF--GTPPLMA-----VVGAIONILSKS 410
DB 323 ATYGSKLKLNISILMSLGFIMLFTTIGLTGIMLSSSIDILHDTYVVGHFHYVLSWG 382
QY 411 TKVALPDS 418
DB 383 AVFAIIS 390
RESULT 36
COXI_PODAN STANDARD; PRT; 541 AA.
AC P20681; O21208; Q35363;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
GN COI OR COI.
OS Podospora anserina.
OG Mitochondrion.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90124722; PubMed=2558809;
RA Cummings D.J., Michel F., McNally K.L.;
RT "DNA sequence analysis of the 24.5 kilobase pair cytochrome oxidase
subunit I mitochondrial gene from Podospora anserina: a gene with
sixteen junctions.";
RL Curr. Genet. 16:381-406(1989).
RN [2]
RP COMPLETE GENOME.
RC STRAIN=s;
RA MEDLINE=90291512; PubMed=2357736;
RA Cummings D.J., McNally K.L., Domencio J.M., Matsura E.T.;
RT "The complete DNA sequence of the mitochondrial genome of Podospora
anserina.";
RL Curr. Genet. 17:375-402(1990).
RN [3]
RP SEQUENCE OF 376-541 FROM N.A.
RC STRAIN=s;
RX MEDLINE=89232730; PubMed=3246349;
RA Viterly-Jamet C.;
RT "Senescence in Podospora anserina: a possible role for nucleic acid

RT interacting proteins suggested by the sequence analysis of a
 RT mitochondrial DNA region specifically amplified in senescent
 RT cultures".
 RL Gene 74:387-398 (1988).
 RN (4)
 RP SEQUENCE OF 4-86 AND 212-242 FROM N.A.
 RX MEDLINE=88223503; PubMed=2836091;
 RA Kueck U., Osiewacz H.D., Schmidt U., Kappelhoff B., Schulte E.,
 RA Strahl U., Esser K.,
 RA "The onset of senescence is affected by DNA rearrangements of a
 RT discontinuous mitochondrial gene in Podospora anserina."
 RL Curr. Genet. 9:373-382 (1985).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 CC c + 2 H(2)O.
 CC -1- PATHWAY: Respiratory chain, terminal step.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X55026; CAA3877.1; -;
 DR EMBL; M28703; AAA32001.2; -;
 DR EMBL; X07119; CAA30131.1; -;
 DR EMBL; X07120; CAA30131.1; JOINED.
 DR EMBL; X07121; CAA30132.1; -;
 DR PIR; A48327; A48327.
 DR HSSP; P98002; IARI.
 DR InterPro; IPR000883; COX1.
 DR Pfam; PF00115; COX1.1.
 DR PRINTS; PR01165; CYCOXIDASRI.
 DR PROSITE; PS00077; COX1.1.
 KM Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 KM Respiratory chain; Inner membrane.
 FT METAL. 68 68 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT METAL. 247 247 COPPER B (PROBABLE).
 FT METAL. 251 251 COPPER B (PROBABLE).
 FT METAL. 296 296 COPPER B (PROBABLE).
 FT METAL. 297 297 COPPER B (PROBABLE).
 FT METAL. 382 382 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).
 FT METAL. 384 384 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT CROSSLINK. 384 384 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT 247 251 N -> T (IN REF. 2 AND 4).
 FT CONFLICT. 23 23 S -> G (IN REF. 2 AND 4).
 FT CONFLICT. 51 51 S -> I (IN REF. 3).
 FT CONFLICT. 414 414 T -> I (IN REF. 3).
 SQ SEQUENCE 541 AA; 59737 MW; 06AF3D497D03DAF2 CRC64;
 Query Match 4.4%; Score 115; DB 1; Length 541;
 Best Local Similarity 20.7%; Pred. No. 2.1; Indels 148; Gaps 20;
 Matches 101; Conservative 60; Mismatches 179;
 QY 51 LIVGAPSGAGBAIPKIFKFWLVVPCAIIMPL-----IYAKLSNLSKQ----- 92
 DB 92 LTVGGPDVAFPRLNINISFWLPPSLILVFSACIEGAGTGWTIPPSPISGVSHSGPSVD 151
 QY 93 -ALF--YAVGTFPLFFALFPVIVPLADV--LHPT--FADRLQAIL--PGILG 139
 DB 152 LAIFALHLSGVSSLLGANNFTITIMNMKTPSRILFKALFGMAVITAVLLLSIPVLAG 211
 QY 140 LVAIL--RNMTPFAFYVLAEIMSGVMSLFWFGPANEITKHEAKR----- 184

DB 212 AITMLTLDRNPN-ISFFETAGGDDPILFQHLFWFGHEVYILLIPAGITSTTSAYSN 270
 QY 185 -----YALFCIGA-----NISLASGAIYASKLRASVSEGVN--DW- 220
 DB 271 KSVFEGYIGMVAAMSGIGLFGYVSHHMYTGLDVTDAFYFAATLLIAVPGIKIFSWL 330
 QY 221 ----GISRLMAMTIVSGVLVMSYMNINQVLTDFPFYVPEENQKKKAKPKPMNKD 276
 DB 331 ATCYGGSIRLTPSMFLFALGFVFMFTGGLSGVILAN-----ASLDIAFHD 375
 QY 277 SFVLYDRSPYILL-----LTLVIAVGINLIEVTWKSQ 312
 DB 376 TVYVAHHVYLSNGAVFAMFSQYFWIPKRLGLNNTLSKVQWILFIGNVTFPFOH 435
 QY 313 KLYQPNM---NDYSEFNGNSFWTGVSVLIMLFVGVGNVIRKFGMLTGAIVTPVM--LL 367
 DB 436 FLGLQGMRRRISDYD-----AFAGWNLISF-----GSITSVAAMWF 474
 QY 368 TGIYFPLV--IFNQASGLVAMFCTPMLAVVGAIONLSTKYLALDSTEMAYI 425
 DB 475 LYIYVLQVBESEYVAGRPFWLNPQFYDTLQ-----ALLNSYPSLEVALSSPPPHAFV 528
 QY 426 PLDQEQKV 433
 DB 529 SLPLQSN 536
 RESULT 37
 COX1_THETH STANDARD; PRT; 562 AA.
 ID COX1_THETH
 AC Q56408;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Cytochrome c ba (3))
 DE subunit I (Cytochrome cba3 subunit 1).
 GN CBA3.
 OS Thermus thermophilus.
 OS Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
 OC Thermus.
 CX NCBI_TaxID=274;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HB8 / ATCC 27634;
 RX MEDLINE=95386472; PubMed=7657607;
 RA Keightley J.A., Zimmermann B.H., Mather M.W., Springer P.,
 RA Pastuszyn A., Lawrence D.M., Fee J.A.;
 RT "Molecular genetic and protein chemical characterization of the
 RT cytochrome ba3 from Thermus thermophilus HB8."
 RT J. Biol. Chem. 270:20345-20358 (1995).
 RN [2]
 RP SEQUENCE OF 336-443 FROM N.A.
 RC STRAIN=VK1;
 RX MEDLINE=97352522; PubMed=9208917;
 RA Ehresmann A., Rak A., Garber M.B., Reinbolt J., Ehresmann B.,
 RA Ehresmann C., Grunberg-Manago M., Portier C.;
 RT "Ribosomal protein S15 from Thermus thermophilus -- cloning,
 RT sequencing, overexpression of the gene and RNA-binding properties of
 RT the protein".
 RT Eur. J. Biochem. 246:291-300 (1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RC STRAIN=HB8 / ATCC 27634;
 RX MEDLINE=20237613; PubMed=10775261;
 RA Soullaine T., Buse G., Bourenkov G.P., Bartunik H.D., Huber R.,
 RA Than M.E.;
 RT "Structure and mechanism of the aberrant ba3-cytochrome c oxidase from
 RT Thermus thermophilus".
 RT EMBO J. 19:1766-1776 (2000).
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 CC c + 2 H(2)O.
 CC -1- COPFACTOR: TWO HEME GROUPS AND COPPER B (BY SIMILARITY).

CC -1- PATHWAY: Respiratory chain; terminal step.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L09121; AAB00370.1; -
DR EMBL: Z84206; CAB06339.1; -
DR PIR: T52481; T52481.
DR PDB: 1EHK; 2I-FEB-01.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1; 1.
DR PRINTS: PR01165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1; 1.
KW Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain;
KW Hydrogen ion transport; 3D-structure.
FT TRANSMEM 21 41 POTENTIAL.
FT TRANSMEM 74 94 POTENTIAL.
FT TRANSMEM 105 125 POTENTIAL.
FT TRANSMEM 144 164 POTENTIAL.
FT TRANSMEM 187 207 POTENTIAL.
FT TRANSMEM 227 247 POTENTIAL.
FT TRANSMEM 267 287 POTENTIAL.
FT TRANSMEM 300 320 POTENTIAL.
FT TRANSMEM 345 365 POTENTIAL.
FT TRANSMEM 385 405 POTENTIAL.
FT TRANSMEM 420 440 POTENTIAL.
FT TRANSMEM 471 491 POTENTIAL.
FT TRANSMEM 527 547 POTENTIAL.
FT METAL 72 72 IRON (HEME A AXIAL LIGAND) (PROBABLE).
FT METAL 233 233 COPPER B (PROBABLE).
FT METAL 237 237 COPPER B (PROBABLE).
FT METAL 282 282 COPPER B (PROBABLE).
FT METAL 283 283 COPPER B (PROBABLE).
FT METAL 384 384 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).
FT METAL 386 386 IRON (HEME A AXIAL LIGAND) (PROBABLE).
FT CROSSLAK 233 237 1'-histidyl-3'-tyrosine.
FT TURN 16 16
FT HELIX 17 45
FT TURN 46 48
FT HELIX 52 58
FT TURN 60 61
FT HELIX 65 77
FT TURN 78 78
FT HELIX 79 97
FT TURN 98 98
FT HELIX 103 125
FT TURN 126 127
FT TURN 133 134
FT TURN 137 138
FT STRAND 139 139
FT HELIX 143 173
FT TURN 175 176
FT HELIX 181 206
FT TURN 207 207
FT TURN 208 212
FT HELIX 213 214
FT STRAND 218 219
FT HELIX 221 232
FT TURN 233 233
FT HELIX 234 249
FT TURN 250 250
FT HELIX 251 254
FT TURN 255 256
FT STRAND 259 259
FT HELIX 262 275
FT TURN 276 279

FT HELIX 280 283
FT TURN 285 286
FT TURN 288 289
FT HELIX 292 305
FT TURN 306 306
FT HELIX 307 326
FT TURN 327 328
FT TURN 332 333
FT HELIX 334 338
FT TURN 341 342
FT HELIX 344 366
FT TURN 367 367
FT HELIX 369 371
FT TURN 372 375
FT HELIX 376 377
FT TURN 378 378
FT TURN 379 379
FT HELIX 380 389
FT TURN 389 391
FT HELIX 392 399
FT HELIX 400 402
FT TURN 403 403
FT HELIX 404 409
FT HELIX 415 444
FT TURN 445 446
FT STRAND 448 448
FT HELIX 453 455
FT TURN 457 458
FT HELIX 461 463
FT HELIX 464 493
FT TURN 502 503
FT STRAND 511 511
FT TURN 515 516
FT HELIX 517 524
FT TURN 525 525
FT HELIX 527 550
FT TURN 551 551
FT STRAND 556 557
FT SEQUENCE 559 559
SQ SEQUENCE 562 AA; 62527 MW; 769B9E2F203617B CRC64;
Query Match 4.4%; Score 115; DB 1; Length 562;
Best Local Similarity 21.0%; Pred. No. 2.2;
Matches 79; Conservative 44; Mismatches 112; Indels 142; Gaps 18;
QY 64 PFKFWLVVPCALIFMLIYAKLSNLSKQALFYAVGTP-----FLIFALFPTV----- 112
DB 234 FIVFWLIPAVAIIV-----TILPKQAGKLVSDPMARLAFLLFLLSTPVGFFHQF 285
QY 113 -----IYPLRDVLAHPT-----EPADRLO-----A 131
DB 286 ADPGIDPFWKMHVSVLTFVAVPESLMTAFYVAASLEFFAGRLGGGLFGWIRALPWNPA 345
QY 132 ILPP--GLGLVA-----ILRN--WTFAPFYV-LAEIWSYMLSLMFWGF 171
DB 346 FVAIVLGLIGITPGAGGIVNASFTLDYVANTAMVPGHFLQVASLVTLLFAMSLVYLL 405
QY 172 ANEITK-IHEAKRYVALFGIANISLASGRIYMAKSLKRAVSVEGVDPWGISRLMMAM 230
DB 406 PNLTKGKPSIDAKR-----RLGLAVVWL-----WFL----- 430
QY 231 TIVSGLVVMA--SYWMINKQVLTDP-RFYNPENKQKKGAKPKNNMKSDFLYLDSPYI 287
DB 431 -----GMMIMAVGLHW--AGILNVPRAYIIAQVEDAIPHAIVP-----MVFNVLAGIV 476
QY 288 LLTLTLVAVGICINLIEVTWKSQK-----LOYPNNDYSEFNGNSFWTGVSV 338
DB 477 LTVALLLFITYGLFVLSRSRKRPDLAPLPFAVISGPEGRRLVLMNDRIGFPAVAIAI 536
QY 339 LITLFPVGGNVIRKFGWL 355
DB 537 LVVLAIGFTLVQVLFQHL 553

RESULT 38
COX1_ACACA STANDARD; PRT; 873 AA.
AC Q37370;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome c oxidase polypeptide I+II (EC 1.9.3.1).
GN COX1/2.
OS Acanthamoeba castellanii (Amoeba).
OG Mitochondrion.
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_Taxid=5755;
RN [1]
RP SEQUENCE FROM N.A. / Neff;
RC STRAIN=ATCC 30010 / Neff;
RX MEDLINE=95147275; PubMed=7844823;
RA Burger G., Plante I., Lonegan K.M., Gray M.W.;
RT "The mitochondrial DNA of the amoeboid protozoan, Acanthamoeba
RT castellanii: complete sequence, gene content and genome
RT organization."
RL J. Mol. Biol. 245:522-537(1995).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO 1 IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.
CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
CC c + 2 H(2)O.
CC -1- PATHWAY: Respiratory chain; terminal step.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE HEME-COPPER
CC RESPIRATORY OXIDASE FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE CYTOCHROME C
CC OXIDASE SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U12386; AAD11820.1; -.
DR PIR: S53828; S53828.
DR HSP: P00396; 20CC.
DR InterPro: IPR001505; Copper_CuA.
DR InterPro: IPR000883; COX1.
DR InterPro: IPR002429; Cyt_c_ox_2.
DR Pfam: PF00115; COX1.1.
DR Pfam: PF00116; COX2.1.
DR Pfam: PF02790; COX2_TM.1.
DR PRINTS: PRO1165; CYCOXIDASE1.
DR PRINTS: PRO1166; CYCOXIDASE1.
DR ProDom: PD000131; Copper_CuA; 1.
DR PROSITE: PS00077; COX1.1.
DR PROSITE: PS00078; COX2.1.
DR Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW Respiratory chain; Inner membrane.
KM DOMAIN 1 474 COX1.
FT DOMAIN 1 475 843 COX2.
FT TRANSMEM 29 49 POTENTIAL.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 91 111 POTENTIAL.
FT TRANSMEM 114 134 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
FT TRANSMEM 197 217 POTENTIAL.
FT TRANSMEM 248 268 POTENTIAL.

FT TRANSMEM 280 300 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 351 371 POTENTIAL.
FT TRANSMEM 390 410 POTENTIAL.
FT TRANSMEM 427 447 POTENTIAL.
FT TRANSMEM 469 489 POTENTIAL.
FT TRANSMEM 536 556 POTENTIAL.
FT TRANSMEM 585 605 POTENTIAL.
FT TRANSMEM 656 676 POTENTIAL.
FT TRANSMEM 845 865 POTENTIAL.
FT TRANSMEM 873 893 POTENTIAL.
FT TRANSMEM 921 941 POTENTIAL.
FT TRANSMEM 978 998 POTENTIAL.
FT TRANSMEM 1015 1035 POTENTIAL.
FT TRANSMEM 1052 1072 POTENTIAL.
FT TRANSMEM 1089 1109 POTENTIAL.
FT TRANSMEM 1126 1146 POTENTIAL.
FT TRANSMEM 1163 1183 POTENTIAL.
FT TRANSMEM 1200 1220 POTENTIAL.
FT TRANSMEM 1237 1257 POTENTIAL.
FT TRANSMEM 1274 1294 POTENTIAL.
FT TRANSMEM 1311 1331 POTENTIAL.
FT TRANSMEM 1348 1368 POTENTIAL.
FT TRANSMEM 1385 1405 POTENTIAL.
FT TRANSMEM 1422 1442 POTENTIAL.
FT TRANSMEM 1459 1479 POTENTIAL.
FT TRANSMEM 1496 1516 POTENTIAL.
FT TRANSMEM 1533 1553 POTENTIAL.
FT TRANSMEM 1570 1590 POTENTIAL.
FT TRANSMEM 1607 1627 POTENTIAL.
FT TRANSMEM 1644 1664 POTENTIAL.
FT TRANSMEM 1681 1701 POTENTIAL.
FT TRANSMEM 1718 1738 POTENTIAL.
FT TRANSMEM 1755 1775 POTENTIAL.
FT TRANSMEM 1792 1812 POTENTIAL.
FT TRANSMEM 1829 1849 POTENTIAL.
FT TRANSMEM 1866 1886 POTENTIAL.
FT TRANSMEM 1903 1923 POTENTIAL.
FT TRANSMEM 1940 1960 POTENTIAL.
FT TRANSMEM 1977 1997 POTENTIAL.
FT TRANSMEM 2014 2034 POTENTIAL.
FT TRANSMEM 2051 2071 POTENTIAL.
FT TRANSMEM 2088 2108 POTENTIAL.
FT TRANSMEM 2125 2145 POTENTIAL.
FT TRANSMEM 2162 2182 POTENTIAL.
FT TRANSMEM 2199 2219 POTENTIAL.
FT TRANSMEM 2236 2256 POTENTIAL.
FT TRANSMEM 2273 2293 POTENTIAL.
FT TRANSMEM 2310 2330 POTENTIAL.
FT TRANSMEM 2347 2367 POTENTIAL.
FT TRANSMEM 2384 2404 POTENTIAL.
FT TRANSMEM 2421 2441 POTENTIAL.
FT TRANSMEM 2458 2478 POTENTIAL.
FT TRANSMEM 2495 2515 POTENTIAL.
FT TRANSMEM 2532 2552 POTENTIAL.
FT TRANSMEM 2569 2589 POTENTIAL.
FT TRANSMEM 2606 2626 POTENTIAL.
FT TRANSMEM 2643 2663 POTENTIAL.
FT TRANSMEM 2680 2700 POTENTIAL.
FT TRANSMEM 2717 2737 POTENTIAL.
FT TRANSMEM 2754 2774 POTENTIAL.
FT TRANSMEM 2791 2811 POTENTIAL.
FT TRANSMEM 2828 2848 POTENTIAL.
FT TRANSMEM 2865 2885 POTENTIAL.
FT TRANSMEM 2902 2922 POTENTIAL.
FT TRANSMEM 2939 2959 POTENTIAL.
FT TRANSMEM 2976 2996 POTENTIAL.
FT TRANSMEM 3013 3033 POTENTIAL.
FT TRANSMEM 3050 3070 POTENTIAL.
FT TRANSMEM 3087 3107 POTENTIAL.
FT TRANSMEM 3124 3144 POTENTIAL.
FT TRANSMEM 3161 3181 POTENTIAL.
FT TRANSMEM 3198 3218 POTENTIAL.
FT TRANSMEM 3235 3255 POTENTIAL.
FT TRANSMEM 3272 3292 POTENTIAL.
FT TRANSMEM 3309 3329 POTENTIAL.
FT TRANSMEM 3346 3366 POTENTIAL.
FT TRANSMEM 3383 3403 POTENTIAL.
FT TRANSMEM 3420 3440 POTENTIAL.
FT TRANSMEM 3457 3477 POTENTIAL.
FT TRANSMEM 3494 3514 POTENTIAL.
FT TRANSMEM 3531 3551 POTENTIAL.
FT TRANSMEM 3568 3588 POTENTIAL.
FT TRANSMEM 3605 3625 POTENTIAL.
FT TRANSMEM 3642 3662 POTENTIAL.
FT TRANSMEM 3679 3699 POTENTIAL.
FT TRANSMEM 3716 3736 POTENTIAL.
FT TRANSMEM 3753 3773 POTENTIAL.
FT TRANSMEM 3790 3810 POTENTIAL.
FT TRANSMEM 3827 3847 POTENTIAL.
FT TRANSMEM 3864 3884 POTENTIAL.
FT TRANSMEM 3901 3921 POTENTIAL.
FT TRANSMEM 3938 3958 POTENTIAL.
FT TRANSMEM 3975 3995 POTENTIAL.
FT TRANSMEM 4012 4032 POTENTIAL.
FT TRANSMEM 4049 4069 POTENTIAL.
FT TRANSMEM 4086 4106 POTENTIAL.
FT TRANSMEM 4123 4143 POTENTIAL.
FT TRANSMEM 4160 4180 POTENTIAL.
FT TRANSMEM 4197 4217 POTENTIAL.
FT TRANSMEM 4234 4254 POTENTIAL.
FT TRANSMEM 4271 4291 POTENTIAL.
FT TRANSMEM 4308 4328 POTENTIAL.
FT TRANSMEM 4345 4365 POTENTIAL.
FT TRANSMEM 4382 4402 POTENTIAL.
FT TRANSMEM 4419 4439 POTENTIAL.
FT TRANSMEM 4456 4476 POTENTIAL.
FT TRANSMEM 4493 4513 POTENTIAL.
FT TRANSMEM 4530 4550 POTENTIAL.
FT TRANSMEM 4567 4587 POTENTIAL.
FT TRANSMEM 4604 4624 POTENTIAL.
FT TRANSMEM 4641 4661 POTENTIAL.
FT TRANSMEM 4678 4698 POTENTIAL.
FT TRANSMEM 4715 4735 POTENTIAL.
FT TRANSMEM 4752 4772 POTENTIAL.
FT TRANSMEM 4789 4809 POTENTIAL.
FT TRANSMEM 4826 4846 POTENTIAL.
FT TRANSMEM 4863 4883 POTENTIAL.
FT TRANSMEM 4900 4920 POTENTIAL.
FT TRANSMEM 4937 4957 POTENTIAL.
FT TRANSMEM 4974 4994 POTENTIAL.
FT TRANSMEM 5011 5031 POTENTIAL.
FT TRANSMEM 5048 5068 POTENTIAL.
FT TRANSMEM 5085 5105 POTENTIAL.
FT TRANSMEM 5122 5142 POTENTIAL.
FT TRANSMEM 5159 5179 POTENTIAL.
FT TRANSMEM 5196 5216 POTENTIAL.
FT TRANSMEM 5233 5253 POTENTIAL.
FT TRANSMEM 5270 5290 POTENTIAL.
FT TRANSMEM 5307 5327 POTENTIAL.
FT TRANSMEM 5344 5364 POTENTIAL.
FT TRANSMEM 5381 5401 POTENTIAL.
FT TRANSMEM 5418 5438 POTENTIAL.
FT TRANSMEM 5455 5475 POTENTIAL.
FT TRANSMEM 5492 5512 POTENTIAL.
FT TRANSMEM 5529 5549 POTENTIAL.
FT TRANSMEM 5566 5586 POTENTIAL.
FT TRANSMEM 5603 5623 POTENTIAL.
FT TRANSMEM 5640 5660 POTENTIAL.
FT TRANSMEM 5677 5697 POTENTIAL.
FT TRANSMEM 5714 5734 POTENTIAL.
FT TRANSMEM 5751 5771 POTENTIAL.
FT TRANSMEM 5788 5808 POTENTIAL.
FT TRANSMEM 5825 5845 POTENTIAL.
FT TRANSMEM 5862 5882 POTENTIAL.
FT TRANSMEM 5899 5919 POTENTIAL.
FT TRANSMEM 5936 5956 POTENTIAL.
FT TRANSMEM 5973 5993 POTENTIAL.
FT TRANSMEM 6010 6030 POTENTIAL.
FT TRANSMEM 6047 6067 POTENTIAL.
FT TRANSMEM 6084 6104 POTENTIAL.
FT TRANSMEM 6121 6141 POTENTIAL.
FT TRANSMEM 6158 6178 POTENTIAL.
FT TRANSMEM 6195 6215 POTENTIAL.
FT TRANSMEM 6232 6252 POTENTIAL.
FT TRANSMEM 6269 6289 POTENTIAL.
FT TRANSMEM 6306 6326 POTENTIAL.
FT TRANSMEM 6343 6363 POTENTIAL.
FT TRANSMEM 6380 6400 POTENTIAL.
FT TRANSMEM 6417 6437 POTENTIAL.
FT TRANSMEM 6454 6474 POTENTIAL.
FT TRANSMEM 6491 6511 POTENTIAL.
FT TRANSMEM 6528 6548 POTENTIAL.
FT TRANSMEM 6565 6585 POTENTIAL.
FT TRANSMEM 6602 6622 POTENTIAL.
FT TRANSMEM 6639 6659 POTENTIAL.
FT TRANSMEM 6676 6696 POTENTIAL.
FT TRANSMEM 6713 6733 POTENTIAL.
FT TRANSMEM 6750 6770 POTENTIAL.
FT TRANSMEM 6787 6807 POTENTIAL.
FT TRANSMEM 6824 6844 POTENTIAL.
FT TRANSMEM 6861 6881 POTENTIAL.
FT TRANSMEM 6898 6918 POTENTIAL.
FT TRANSMEM 6935 6955 POTENTIAL.
FT TRANSMEM 6972 6992 POTENTIAL.
FT TRANSMEM 7009 7029 POTENTIAL.
FT TRANSMEM 7046 7066 POTENTIAL.
FT TRANSMEM 7083 7103 POTENTIAL.
FT TRANSMEM 7120 7140 POTENTIAL.
FT TRANSMEM 7157 7177 POTENTIAL.
FT TRANSMEM 7194 7214 POTENTIAL.
FT TRANSMEM 7231 7251 POTENTIAL.
FT TRANSMEM 7268 7288 POTENTIAL.
FT TRANSMEM 7305 7325 POTENTIAL.
FT TRANSMEM 7342 7362 POTENTIAL.
FT TRANSMEM 7379 7399 POTENTIAL.
FT TRANSMEM 7416 7436 POTENTIAL.
FT TRANSMEM 7453 7473 POTENTIAL.
FT TRANSMEM 7490 7510 POTENTIAL.
FT TRANSMEM 7527 7547 POTENTIAL.
FT TRANSMEM 7564 7584 POTENTIAL.
FT TRANSMEM 7601 7621 POTENTIAL.
FT TRANSMEM 7638 7658 POTENTIAL.
FT TRANSMEM 7675 7695 POTENTIAL.
FT TRANSMEM 7712 7732 POTENTIAL.
FT TRANSMEM 7749 7769 POTENTIAL.
FT TRANSMEM 7786 7806 POTENTIAL.
FT TRANSMEM 7823 7843 POTENTIAL.
FT TRANSMEM 7860 7880 POTENTIAL.
FT TRANSMEM 7897 7917 POTENTIAL.
FT TRANSMEM 7934 7954 POTENTIAL.
FT TRANSMEM 7971 7991 POTENTIAL.
FT TRANSMEM 8008 8028 POTENTIAL.
FT TRANSMEM 8045 8065 POTENTIAL.
FT TRANSMEM 8082 8102 POTENTIAL.
FT TRANSMEM 8119 8139 POTENTIAL.
FT TRANSMEM 8156 8176 POTENTIAL.
FT TRANSMEM 8193 8213 POTENTIAL.
FT TRANSMEM 8230 8250 POTENTIAL.
FT TRANSMEM 8267 8287 POTENTIAL.
FT TRANSMEM 8304 8324 POTENTIAL.
FT TRANSMEM 8341 8361 POTENTIAL.
FT TRANSMEM 8378 8398 POTENTIAL.
FT TRANSMEM 8415 8435 POTENTIAL.
FT TRANSMEM 8452 8472 POTENTIAL.
FT TRANSMEM 8489 8509 POTENTIAL.
FT TRANSMEM 8526 8546 POTENTIAL.
FT TRANSMEM 8563 8583 POTENTIAL.
FT TRANSMEM 8600 8620 POTENTIAL.
FT TRANSMEM 8637 8657 POTENTIAL.
FT TRANSMEM 8674 8694 POTENTIAL.
FT TRANSMEM 8711 8731 POTENTIAL.
FT TRANSMEM 8748 8768 POTENTIAL.
FT TRANSMEM 8785 8805 POTENTIAL.
FT TRANSMEM 8822 8842 POTENTIAL.
FT TRANSMEM 8859 8879 POTENTIAL.
FT TRANSMEM 8896 8916 POTENTIAL.
FT TRANSMEM 8933 8953 POTENTIAL.
FT TRANSMEM 8970 8990 POTENTIAL.
FT TRANSMEM 9007 9027 POTENTIAL.
FT TRANSMEM 9044 9064 POTENTIAL.
FT TRANSMEM 9081 9101 POTENTIAL.
FT TRANSMEM 9118 9138 POTENTIAL.
FT TRANSMEM 9155 9175 POTENTIAL.
FT TRANSMEM 9192 9212 POTENTIAL.
FT TRANSMEM 9229 9249 POTENTIAL.
FT TRANSMEM 9266 9286 POTENTIAL.
FT TRANSMEM 9303 9323 POTENTIAL.
FT TRANSMEM 9340 9360 POTENTIAL.
FT TRANSMEM 9377 9397 POTENTIAL.
FT TRANSMEM 9414 9434 POTENTIAL.
FT TRANSMEM 9451 9471 POTENTIAL.
FT TRANSMEM 9488 9508 POTENTIAL.
FT TRANSMEM 9525 9545 POTENTIAL.
FT TRANSMEM 9562 9582 POTENTIAL.
FT TRANSMEM 9599 9619 POTENTIAL.
FT TRANSMEM 9636 9656 POTENTIAL.
FT TRANSMEM 9673 9693 POTENTIAL.
FT TRANSMEM 9710 9730 POTENTIAL.
FT TRANSMEM 9747 9767 POTENTIAL.
FT TRANSMEM 9784 9804 POTENTIAL.
FT TRANSMEM 9821 9841 POTENTIAL.
FT TRANSMEM 9858 9878 POTENTIAL.
FT TRANSMEM 9895 9915 POTENTIAL.
FT TRANSMEM 9932 9952 POTENTIAL.
FT TRANSMEM 9969 9989 POTENTIAL.
FT TRANSMEM 10000 10000 POTENTIAL.

GN MTCYB OR COB OR CYTB.
 OS Pinguinus impennis (Great auk).
 CC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Charadriiformes; Laridae; Pinguinus.
 OX NCBI_TaxID=94623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Feather, and Skin;
 RX MEDLINE=22189004; PubMed=12200471;
 RA Moun T., Arnason U., Arnason B.;
 RT "Mitochondrial DNA sequence evolution and phylogeny of the Atlantic
 Alcididae, including the extinct great auk (*Pinguinus impennis*).";
 RL Mol. Biol. Evol. 19:1434-1439(2002).
 CC -1- FUNCTION: Component of the ubiquinol-cytochrome c reductase
 complex (complex III or cytochrome b-c1 complex), which is a
 respiratory chain that generates an electrochemical potential
 coupled to ATP synthesis (By similarity).
 CC -1- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
 b562) is low-potential and absorbs at about 562, and heme 2 (or BH
 or b566) is high-potential and absorbs at about 566 (By
 similarity).
 CC -1- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
 cytochrome c1 and the Rieske protein (By similarity).
 CC -1- SIMILARITY: Belongs to the cytochrome b family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ242685; CAC80349.1; -
 DR InterPro; IPR005798; Cytb b6_C.
 DR InterPro; IPR005797; Cytb b6_N.
 DR Pfam; PF00032; cytochrome_b_C; 1.
 DR Pfam; PF00033; cytochrome_b_N; 1.
 DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
 DR PROSITE; PS00193; CYTOCHROME_B_QQ; 1.
 DR Electon transport; Mitochondrion; Respiratory chain; Transmembrane;
 Heme.
 KM METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 183 183 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
 SQ SEQUENCE 380 AA; 42549 MW; A830ADC6E8FDDC2 CRC64;
 Query Match 4.4%; Score 114.5; DB 1; Length 380;
 Best Local Similarity 20.8%; Pred. No. 1.6; Indels 87; Gaps 14;
 Matches 64; Conservative 54; Mismatches 103;
 QY 136 GLLGLVILTRMTFAFYALAEWGSWVLSMFMGPAEITKIHNA----- 181
 DB 117 GIILLITLMA--TAVGVLP--WG-----QMSFNG-AVITNLSAITYIGQLVEMWG 167
 QY 182 -----KRPYALFGIGANISLLASGRAIVASKLRAVSEGVDPNGI----- 222
 DB 168 GFSVDNPTLTPFPALHFL--LPFWIAGIALIHILFTLHESGSN--NPGLISNCDKIPFH 222
 QY 223 ---SLRLMAMTIYSGVLVMSYMWINKNVLTDPFVYDEMGKKKAKPRMNMKDSFL 279
 DB 223 PYFSIKDILGFTIM--FLPFLTLTALFSPNLLGDPENFNPANPLVTPPHIKBWFLEFAY 280
 QY 280 YLDRSPYLLTLTLVIAIGICINILIEVTKSQKLQYNNMNDYSEFMNGSFFMTGVSVL 339
 DB 281 ILRSEIPNLTGGVIALAAVIVLPLVPLHKS--KQRATFRPLSGFL---FWITIVANLL 334
 QY 340 IMLFVGGNVIRKFGMLTALVTPVWVL--LTGIVFPALVIFRNQASGLVAMFGTTPMLA 397
 DB 335 ILTWVGSQPVH-----PFIILIGLASLIVFTIL-----ILLF 367

QY 398 VVVGAIQN 405
 DB 368 PIIQALEN 375
 RESULT 40
 Y917_MYCTU STANDARD; PRT; 593 AA.
 AC 005909;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable transport protein RV0917.
 GN RV0917 OR MT0942 OR MTCY21C12.11.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=9634230; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby R., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ernolova M.D., Salzberg S.L.,
 RA DeLor A., Ulteback T., Weisman J., Knout H., Gill J., Mikula A.,
 RA Bhalal W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE BCCT (TC 2.A.15) FAMILY OF
 TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z95210; CAB08499.1; -
 DR EMBL; AE006980; AAK45189.1; -
 DR PIR; E70582; E70582.
 DR TIGR; MT0942; -
 DR Tuberculat; RV0917; -
 DR InterPro; IPR000060; BCCT_transporter.
 DR Pfam; PF02028; BCCT_1.
 DR Pfam; PF02028; BCCT_1.
 DR PROSITE; PS01303; BCCT_1.
 DR PROSITE; PS01303; BCCT_1.
 KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
 FT TRANSMEM 21 41 POTENTIAL.
 FT TRANSMEM 60 80 POTENTIAL.
 FT TRANSMEM 97 117 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT TRANSMEM 204 224 POTENTIAL.
 FT TRANSMEM 243 263 POTENTIAL.

THIS PAGE BLANK (USPTO)

Db 211280 CAGCTTCGGCTGCTCGCTATGTTGGTACACCTCTCTCATGCTAGCTGTTGTC 211221
Qy 401 GYAlaAlaIleGlnAsnIleuSerLysSerThrLysValAlaLeuPheAspSerThrLys 420
Db 211220 GAGACATACAGAAATTTCTTTCGAAATCCAAATAACGCTCTCTTGACTCACTATA 211161
Qy 421 GluMetCAlaTyrIleProLeuAspGlnGlnIleValIleGlyLysValAlaIleAsp 440
Db 211160 GAAATGCCCTATATCCCTCTTGAACAGACAAAAGTCMAAGTAAGCTGCTATTGAT 211101
Qy 441 ValValAlaAlaIleArgPheGlyLysSerGlyValAlaLeuIleGlnGlnIleLeuVal 460
Db 211100 GTAGTTCCTCCCGCTTCGAAATACGAGAGCTTATTCACAAAGCTTGGCTGCT 211041
Qy 461 IleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValIleLeuLeuPheIle 480
Db 211040 ATCTGCGAGATATGAGATAGACCCCTTATCTTGCAGTGAATCTTCTTTCATCAT 210981
Qy 481 AlaIleTyrLeuValSerAlaThrLysLeuAsnLysLeuPheLeuIleGlnIleAlaLeu 500
Db 210980 GCTATTGTTGTTGTTCTGCACTAAGTAAACAACTATCTTGAAGCAGCTGCTCTT 210921
Qy 501 LysGlnGlnGlnValAlaIleGlnGlnAspSerAlaProIleSer 515
Db 210920 AAAGACAGAGAGTGGCTCAAGAAATTCAGCTCTCTCTTCA 210876

RESULT 3
AAK91990
ID AAK91990 standard; DNA; 1230025 BP.

AAK91990;
13-SEP-1999 (first entry)

Nucleotide sequence of the complete genome of Chlamydia pneumoniae.

Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
vaccine; neutralising epitope; ss.

Chlamydia pneumoniae.

W09927105-A2.
03-JUN-1999.

20-NOV-1998; 98WC-1B01890.

04-NOV-1998; 98US-0107078.

21-NOV-1997; 97FR-0014673.

(GEST) GENSET.

Griffiths R;

WPI; 1999-357842/30.

Genome sequence of Chlamydia pneumoniae

Claim 1; Page 291-611; 1912p; English.

The present sequence represents the complete genome of Chlamydia pneumoniae, and encodes proteins AY34584-Y35879. C. pneumoniae causes respiratory disease such as pneumonia and bronchitis, and is thought to be a contributing factor in heart disease, sarcoidosis, sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The polypeptides encoded by the open reading frames of the C. pneumoniae genome (see AY34584-Y35879) can be used in immunogenic compositions as vaccines. Vectors containing C. pneumoniae nucleotide sequences can also be used as immunogenic compositions, especially where the vector directs the expression of a neutralising epitope of C. pneumoniae.

Griffiths
Nucleic acid sequencing primer
102 (c)

Alignment Scores:
Pred. No.: 7,61e-271 Length: 1230025
Score: 2617.00 Matches: 513
Percent Similarity: 99.61% Conservative: 0
Best Local Similarity: 99.61% Mismatches: 2
Query Match: 99.51% Indels: 0
DB: 20 Gaps: 0

US-09-869-433-2 (1-515) x AAK91990 (1-1230025)

Qy 1 MetThrLysThrGlnGlnLysProPheGlyLysLeuArgSerPheLeuThrProIleHis 20
Db 404370 ATGACAAAAACCGAAGAAAAACCTTTGAAATATGCGCTCTTTGGCGGATCAAT 404429
Qy 21 ThrHisGluLeuLysValLeuPheMetPheLeuMetPhePheCysIleThrPheAsn 40
Db 404430 ACTCAGAGCTTAAAGAAAGTTCTGCAATGTTCCAAATGTTCTTGATTAATCAATTAC 404489
Qy 41 TyrThrValLeuArgAspThrLysAspThrLeuIleValGlyAlaProGlySerGlyAla 60
Db 404490 TATTACGGTGTACGGGATACAAAGACACTTATATGGGAGCTTCGGTTCGGTGA 404549
Qy 61 GluAlaIleProPheIleLysPheThrLeuValAlaProCysAlaIleIlePheMetLeu 80
Db 404550 GAGGCATATCCTTTCATCAAGTTTGGCTTGTGCTTGTGCTTGTGCTTGTGCTT 404609
Qy 81 IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr 100
Db 404610 ATTATGCAAGAGTATGATATTTTAAAGACAGAGCTTATTTATGACAGTGGAGAG 404669
Qy 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
Db 404670 CCTTTTAAATTTCTTGGCCCTGTCGCCAGCTGAATTAATTCGCTACGCGAGTTTA 404729
Qy 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProGlyLeuLeuGlyLeu 140
Db 404730 CATCTTCAAGAAATTTGTCACCTTACAGGCACTTACCTCAGAGTTGCTAGAGATC 404789
Qy 141 ValAlaIleLeuArgAsnThrPheAlaAlaPheTyrValLeuAlaGluLeuTyrGly 160
Db 404790 GTTGCATCTTAAAGAACTGACATTTGCGATTTATGATCTGTAACATATGGGA 404849
Qy 161 SerValMetLeuSerLeuMetPheTyrGlyPheAlaAsnGluIleThrLysIleHisGlu 180
Db 404850 AGCGTATGCTATCTTAATGTTCTGGGAAATTCCTATGAAATTAACAAATCAAGAA 404909
Qy 181 AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
Db 404910 GCAAGAGGTTCTACGCTCTTTTGGATGAGAGCTTAATTTCTTTACAGCTTCGT 404969
Qy 201 ArgAlaIleValIleTyrAlaSerLysLeuArgAlaSerValSerGlyValAspProTyr 220
Db 404970 CGTGCATTTGTTGGGCTTCAAAATGAGAGCTTCGGTTCTGAAAGGTGATCCCTGG 405029
Qy 221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla 240
Db 405030 GGAATTTCTTACGCTTTTATGATGCTATGATGATGATGATGATGATGATGATGAT 405089
Qy 241 SerTyrTyrPheIleAsnLysValLeuThrAspProCysPheTyrAspProGlnGlu 260
Db 405090 AGTTACTGTGTGATCAATTAAGAGGATATGACCGATCTCGCTTCTATATCCAGAAAG 405149
Qy 261 MetGlnLysGlyLysLeuGlyValAlaLysProLysMetAsnMetLysAspSerPheLeuTyr 280
Db 405150 ATGCAAAAGGAGAAAGGCTTAACCTAAAGAAATGAAATGAAATGAAATGAAATG 405209
Qy 281 LeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyLysCys 300
Db 405210 CTTCGTAAGTATCCCTTAATATCTTTATTAAGTCTCTGCTGATTAAGCTGATTAAG 405269
Qy 301 IleLeuLeuIleGlnValThrTyrLysSerGlnLysLeuGlnTyrProAsnMetAsn 320

Db 405270 ATTAACCTTAATCGAAGTACTTGAGAAAAGTCAGCTGAACCTGCAATATCTTAATATGAAT 405329
Qy 321 AsPTySergIuphMetGlyAsnPheserPheTrpThrglyValaIseValleuIle 340
Db 405330 GACTATAGTAGTTCATCGGGAACCTTCCTTCGAGATGGCCGATATCCGTAATC 405389
Qy 341 MetLeuPheValGlyValAsnValIleArgGlyPheGlyYrPLeuThrglyValaIseVal 360
Db 405390 ATGCTATTTGTTGGTGGTAAAGCTATTCGTAATTTGGATGGTAACTGGAGCCCTAGTC 405449
Qy 361 ThrProValMetValleuLeuThrglyYleValaPhePheValaIseValIlePheArgAsn 380
Db 405450 ACTCTCTCATGCTTCTCTTCAACGATTCGTTTCTTCGCTCTGTATCTTTAGAAC 405509
Qy 381 GlnAlaSerGlyLeuValaIaMetPheGlyThrThrProLeuMetLeuAlaValaIseVal 400
Db 405510 CAAGCTTCTGGGCTGGTGGTATCTTGGTAACTCTCTCATGCTGATGCTGCTGTC 405569
Qy 401 GlyAlaIleGlnAsnIleLeuSerLysSerThrLysTrpAlaLeuPheAspSerThrLys 420
Db 405570 GGAGCTATACAGATATCTTTCGAAATTCACAAATAACGCTCTTGACTCACTAA 405629
Qy 421 GlnMetAlaTrpIleProLeuAspGlnGluGlnValaIseValaIleAsp 440
Db 405630 GAAATGGCTATATCCTCTTACCAAGCAAGCAAAAGTCAAGGTAAGCTGCTATTGAT 405689
Qy 441 ValValaIaAlaArgPheGlyLysSerGlyValaLeuIleGlnGlnIleLeuVal 460
Db 405690 GTAGTTCGCCCGCTTCGGAATAATCAGAGAGCTTAACTCAACAAGCTTCTGCTT 405749
Qy 461 IleCysGlySerIleGlyAlaMetThrProTrpLeuAlaValIleLeuPheIleIle 480
Db 405750 ATCTGGAGATATGAGCTATGACCCCTTAATCTTCAGTGAATCTTTTCATCAT 405809
Qy 481 AlaIleTrpLeuValSerAlaTrpLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeu 500
Db 405810 GCTATTTGCTGCTTCTTCTGCACTAACTTAACAACACTATTTCTAGCGAGCTGCTT 405869
Qy 501 LysGlnGlnIleValaIaGlnIleAspSerAlaProAlaSerSer 515
Db 405870 AAGAACAAAGAGTGGCTCAAGAAGATTGAGCTCTCTCTTCA 405914
RESULT 4
AA201425
ID AA201425 standard; DNA; 1038602 BP.
AC AA201425;
XX
DT 07-OCT-1999 (first entry)
XX
DE Complete genome sequence of Chlamydia trachomatis.
XX
KM Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KM paratrachoma; inclusion conjunctivitis; genital disease; perilepetic;
KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KM bartolinitis; pneumopathy; venereal lymphogranulomatosis; ss.
OS Chlamydia trachomatis.
XX
PN MO9928475-A2.
XX
PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98MO-IB01939.
XX
PR 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
PA (GEST) GENSET.
XX
PI Griffeats R;
XX

DR WPI; 1999-371125/31.
XX
PT Genome sequence of Chlamydia trachomatis
XX
PS Claim 1, Page 373-656; 1755pp; English.
XX
CC The present sequence represents the complete genome of Chlamydia
CC trachomatis. Open reading frames (ORFs) of the genome encode
CC polypeptides AY36754-Y37949. The polypeptides can be used as vaccines
CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
CC be used to control growth of the microorganism. Chlamydia trachomatis is
CC responsible for a large number of diseases, e.g. eye diseases such as
CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
CC conjunctivitis; genital diseases such as nongonococcal urethritis;
CC epididymitis, cervicitis, salpingitis, perinephritis, bartolinitis;
CC pneumopathy in breast feeding infants; and venereal
CC lymphogranulomatosis. The polypeptides of the invention may be of use in
CC treating these diseases.
XX
SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;
XX
Alignment Scores:
Pred. No.: 4.77e-219 Length: 1038602
Score: 2141.00 Matches: 409
Percent Similarity: 87.81% Conservative: 52
Best Local Similarity: 77.90% Mismatches: 50
Query Match: 81.41% Indels: 14
DB: 20 Gaps: 3
US-09-869-433-2 (1-515) x AA201425 (1-1038602)
Qy 1 MetThrLysThrGlnGluLysProPheGlyLysLeuArgSerPheLeuTrpProIleHis 20
Db 392407 ATGACTCAAAACCGCGAATAACCTTTGAAAATTCGCTCTTCTTGGCCGATAC 392466
Qy 21 ThrHISGluLeuLysValLeuProMetPheLeuMetPhePheCysIleThrPheAsn 40
Db 392467 ATGCATGAGCTGAGAAAGTTTGCACATGTTCCATGCTTCTTCTGATTTTCATTCAT 392526
Qy 41 TyrThrValleuLysAspThrLysAspThrLeuIleValaIleProGlySerGlyAla 60
Db 392527 TACAGATTTTGAAGATACAAAGATCTTATCTGTTACAGACCGGATCTGGAGCA 392586
Qy 61 GlnAlaIleProPheIleLysPheTrpLeuValaIleProCysAlaIleLeuPheIleLeu 80
Db 392587 GAGGCCATTTCTTTATTAAGTTGCTGCTAGTCGCTCTTCTGCTGTTGTTCAATGCTG 392646
Qy 81 IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaIleGlyThr 100
Db 392647 ATCTACGCCAAGCTAAGCAATTTTGAACAACAGGCTCTTTCTTCCAGCTGCTCTCA 392706
Qy 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
Db 392707 CCATTCGTTGATTTCTTGGCACTGTTCCCTGGGATCTACCCCTTGCATATTTCTT 392766
Qy 121 HisProThrGluPheAlaAspArgGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140
Db 392767 CACCCAAAGCTTGGCTGATACCTTACAAATCATTTCTTCCGGAATTAATGAGATTC 392826
Qy 141 ValAlaIleLeuArgAsnTrpThrPheAlaIaIaPheTyrValaIleGluLeuTrpGly 160
Db 392827 ATTCGATGCTAAGCAATGAGCAATTTGCTGTTCTTCAAGCTTTCGAACTTGGGGA 392886
Qy 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleHisGlu 180
Db 392887 AGCGTTATGCTCTCTTATATCTTCTGGGCTTTCGCAATGAATTAATAATTAAGCCAA 392946
Qy 181 AlaLysArgPheTyrAlaLeuPheGlyYleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
Db 392947 GCTAAAGCTTCTTACGCTCTTTTGGAGATGAGACTTAACGATGCTCTTAAATTTCTGCT 393006
Qy 201 ArgAlaIleValaIleTrpAlaSerLysLeuArgAlaSerValSerGlnGlyValaIleProTrp 220

CC frames in the complete genome (see AAY3153) of *Chlamydia pneumoniae*.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY3153-135879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.

XX Sequence 524 AA;

XX Query Match 99.5%; Score 2617; DB 20; Length 524;

XX Best Local Similarity 99.6%; Pred. No. 6,2e-281;

XX Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTKTEKPGKLRSGFLMPLHTEHKKVLPWLMFCITFNVTYLRDRTDILVAGPSSGA 60
 DB 10 MTKTEKPGKLRSGFLMPLHTEHKKVLPWLMFCITFNVTYLRDRTDILVAGPSSGA 69
 QY 61 EAIPIKFWLVVPCALIFMLIYAKLSNLSKQALFVAVGTPLIFPALFPVITYPLRDVL 120
 DB 70 EAIPIKFWLVVPCALIFMLIYAKLSNLSKQALFVAVGTPLIFPALFPVITYPLRDVL 129
 QY 121 HPTFADRLQALIPGLGLVAIIRNMTFAAFVYLAELGWSVLSLMEFGPANEITKTHE 180
 DB 130 HPTFADRLQALIPGLGLVAIIRNMTFAAFVYLAELGWSVLSLMEFGPANEITKTHE 189
 QY 181 AKRFYALFGIGANISLASGRAIIVASKLRASVSEGVDPWGISLRLLMAMTIVSGLVMA 240
 DB 190 AKRFYALFGIGANISLASGRAIIVASKLRASVSEGVDPWGISLRLLMAMTIVSGLVMA 249
 QY 241 SYWMINKVLTDRFYNPEEMOKGKGAKPMMKDSFYLDSPYITLLTLVAVGIC 300
 DB 250 SYWMINKVLTDRFYNPEEMOKGKGAKPMMKDSFYLDSPYITLLTLVAVGIC 309
 QY 301 INLEIETWKSQKLQYNNNDYSEFMGNFSFWTGVSVLIMLFVGQNVIRKFGMLTGALV 360
 DB 310 INLEIETWKSQKLQYNNNDYSEFMGNFSFWTGVSVLIMLFVGQNVIRKFGMLTGALV 369
 QY 361 TPVWVLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKALPDSRK 420
 DB 370 TPVWVLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKALPDSRK 429
 QY 421 EMAYIPLDOEKVKGAIDVVAARFGKSGGALLIOGGLVTCGSGAMTPYLAIVLFTI 480
 DB 430 EMAYIPLDOEKVKGAIDVVAARFGKSGGALLIOGGLVTCGSGAMTPYLAIVLFTI 489
 QY 481 AIMVLSATKLNKFLAQSALKEQEVVAOEDSAPASS 515
 DB 490 AIMVLSATKLNKFLAQSALKEQEVVAOEDSAPASS 524

RESULT 3
 AAY3153 standard; Protein; 529 AA.

AC AAY3153;

DT 07-OCT-1999 (first entry)

DE Chlamydia trachomatis transport protein.

KM Vaccines; eye disease; conventional trachoma; nonendemic trachoma;
 KM paratrachoma; inclusion conjunctivitis; genital disease; perhepatitis;
 KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KM bartonellosis; pneumonia; venereal lymphogranulomatosis.

XX Chlamydia trachomatis.

OS
 XX
 PN WO9928475-A2.

PD 10-JUN-1999.
 XX
 PF 27-NOV-1996; 96WO-1B01939.
 XX
 PR 04-NOV-1998; 98US-0107077.
 PR 28-NOV-1997; 97ER-0015041.
 PR 17-DEC-1997; 97FR-0016034.
 XX
 PA (BEST) GENSET.
 XX
 PI Griffiths R;
 DR WPI, 1999-371125/31.

PI Genome sequence of *Chlamydia trachomatis*

PS Disclosure; Page 935-936; 1755pp; English.

CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 CC of *Chlamydia trachomatis* (see AAY3153). The polypeptides can be used as
 CC vaccines against *Chlamydia trachomatis*. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. *Chlamydia*
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conjunctivitis, trachoma, nonendemic trachoma,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perhepatitis, bartonellosis; pneumonia; venereal lymphogranulomatosis;
 CC may be of use in treating these diseases.

SQ Sequence 529 AA;

Query Match 81.4%; Score 2141; DB 20; Length 529;

Best Local Similarity 77.9%; Pred. No. 4e-228;

Matches 409; Conservative 52; Mismatches 50; Indels 14; Gaps 3;

QY 1 MTKTEKPGKLRSGFLMPLHTEHKKVLPWLMFCITFNVTYLRDRTDILVAGPSSGA 60
 DB 2 MTKTEKPGKLRSGFLMPLHTEHKKVLPWLMFCITFNVTYLRDRTDILVAGPSSGA 61
 QY 61 EAIPIKFWLVVPCALIFMLIYAKLSNLSKQALFVAVGTPLIFPALFPVITYPLRDVL 120
 DB 62 EAIPIKFWLVVPCALIFMLIYAKLSNLSKQALFVAVGTPLIFPALFPVITYPLRDVL 121
 QY 121 HPTFADRLQALIPGLGLVAIIRNMTFAAFVYLAELGWSVLSLMEFGPANEITKTHE 180
 DB 122 HPTFADRLQALIPGLGLVAIIRNMTFAAFVYLAELGWSVLSLMEFGPANEITKTHE 181
 QY 181 AKRFYALFGIGANISLASGRAIIVASKLRASVSEGVDPWGISLRLLMAMTIVSGLVMA 240
 DB 182 AKRFYALFGIGANISLASGRAIIVASKLRASVSEGVDPWGISLRLLMAMTIVSGLVMA 241
 QY 241 SYWMINKVLTDRFYNPEEMOKGKGAKPMMKDSFYLDSPYITLLTLVAVGIC 300
 DB 242 SYWMINKVLTDRFYNPEEMOKGKGAKPMMKDSFYLDSPYITLLTLVAVGIC 309
 QY 301 INLEIETWKSQKLQYNNNDYSEFMGNFSFWTGVSVLIMLFVGQNVIRKFGMLTGALV 360
 DB 300 INLEIETWKSQKLQYNNNDYSEFMGNFSFWTGVSVLIMLFVGQNVIRKFGMLTGALV 369
 QY 361 TPVWVLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKALPDSRK 420
 DB 360 TPVWVLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKALPDSRK 419
 QY 421 EMAYIPLDOEKVKGAIDVVAARFGKSGGALLIOGGLVTCGSGAMTPYLAIVLFTI 480
 DB 420 EMAYIPLDOEKVKGAIDVVAARFGKSGGALLIOGGLVTCGSGAMTPYLAIVLFTI 479
 QY 481 AIMVLSATKLNKFLAQSALKEQEVVAOEDSAPASS 513
 DB 480 AIMVLSATKLNKFLAQSALKEQEVVAOEDSAPASS 524

11

12

OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (885001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (900001)..(915000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature

Alignment Scores:
Pred. No.: 4,54e-298 Length: 1230025
Score: 2617.00 Matches: 513
Percent Similarity: 99.61% Conservative: 0
Best Local Similarity: 99.61% Mismatches: 2
Query Match: 99.51% Indels: 0
Gaps: 0

seq. ID no. 102 (e)

seq. ID no. 102

seq. ID no. 1

US Patent 6,559,924

US Patent 6,559,924

US-09-869-433-2 (1-515) x US-09-198-452A-1 (1-1230025)
QY 1 MetThyrSthrGluGluValProPheGlyValLeuArgSerPheLeuTrpProIleHis 20
DB 404370 ATGCAAAACCGAAGAAACCTTTGGAAATTTGGCCTTTCTTGGCCGATACAT 404429
QY 21 ThrHsgLueuValValLeuProMetPheLeuMetPhePheCysIleThrPheAsn 40
DB 404430 ACTCAGCGCTTAAGAAAGCTTCCGCAATGTTCCCAATGTTCTTGTATTAACAATTAA 404489
QY 41 TyrThrValLeuArgAspThrLysAspThrLeuIleValGlyValProGlySerGlyValA 60
DB 404490 TATACGCGTTCGATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 404549
QY 61 GluAlaIleProPheIleLysPheTrpLeuValValProCysAlaIleIlePheMetLeu 80
DB 404550 GAGGCAATACCTTTCATCAAGTTTGGCTTGTTCCTTGGCTTTCATTAATTAATTAATTAATTA 404609
QY 81 IleThrValLeuArgSerAsnIleLeuSerLysGlnAlaLeuPheThrValValGlyThr 100
DB 404610 ATTATGCAAGCTAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 404669
QY 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleThrProLeuArgAspValLeu 120
DB 404670 CCCCTTTTAATTTCTTTCCTGCTTCCGCTTCCGCTTATTAATTAATTAATTAATTAATTAATTA 404729
QY 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140
DB 404730 CATCTCAAGAAATTTGGCTTTCATCAAGTTTGGCTTTCATTAATTAATTAATTAATTAATTAATTA 404789
QY 141 ValAlaIleLeuArgAsnTrpThrPheAlaIlePheThrValValLeuAlaGluLeuTrpGly 160
DB 404790 GTTGCATCTTAAGAACTGACATTTCTGCTATTTTAATTAATTAATTAATTAATTAATTAATTAATTA 404849
QY 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleHisGlu 180
DB 404850 AGCGTATGCTATCTTAATTTCTGCGGATTTGCTTAATTAATTAATTAATTAATTAATTAATTAATTA 404909
QY 181 AlaLysArgPheThrValLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
DB 404910 GCAAGCGTTTCTACGCTTTTCCGTTATCGAGCTTAATTAATTAATTAATTAATTAATTAATTAATTA 404969
QY 201 ArgAlaIleValTrpAlaSerLysLeuArgAlaSerValSerGluGlyValAspProTrp 220

DB 404970 CGTGCAATTTGTTGGCTTCAAGTTGAGAGCTTCGTTCTGAAAGTTGATGCTTGG 405029
QY 221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla 240
DB 405030 GGAATTTCTTACGCTTTTGAAGGCTATGATGATGATGATGATGATGATGATGATGATGATG 405089
QY 241 SerTrpTrpTrpIleAsnLysAsnValLeuThrAspProArgPheThrAsnProGlu 260
DB 405090 AGTTACGCTGATCAATTAAGAGCTATGACCGATTCCTGCTTCTTAATTAATTAATTAATTAATTA 405149
QY 261 MetGluValGlyLysValGlyAlaLysProLysMetAsnMetLysAspSerPheLeuTrp 280
DB 405150 ATGCAAAAGGAGAAAGAGTCTTAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 405209
QY 281 LeuAspArgSerProTrpIleLeuLeuLeuThrLeuValIleAlaTrpGlyIleCys 300
DB 405210 CTTCGATATCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 405269
QY 301 IleAsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGlnTrpProAsnMetAsn 320
DB 405270 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 405329
QY 321 AspTrpSerGluPheMetGlyAsnPheSerPheTrpThrGlyValValSerValLeuIle 340
DB 405330 GACTATGAGTCAATGAGGAACTTCTCTGCTGAGCTGGCGTATGCTTCTTATTC 405389
QY 341 MetLeuPheValGlyValAsnValIleArgLysPheGlyTrpLeuThrGlyValLeuVal 360
DB 405390 ATGCTATTTGTTGGTGAACCTCATCGTAATTTGATGATGATGATGATGATGATGATGATGATG 405449
QY 361 ThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn 380
DB 405450 ACTCCGTGATGCTTCTTCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 405509
QY 381 GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValValVal 400
DB 405510 CAAGCTTCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 405569
QY 401 GlyAlaIleGlnAsnIleLeuSerLysSerThrLysThrValLeuPheAspSerThrLys 420
DB 405570 GGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 405629
QY 421 GluMetAlaTrpIleProLeuAspGlnGlnLysValLysGlyValAlaIleAsp 440
DB 405630 GAAATGCGCTATACCTCTTGGACCAAGCAAGAAAGTCAAGATTAAGTGGCTTAT 405689
QY 441 ValValAlaAlaArgPheGlyLysSerGlyValAlaLeuIleGlnGlnGlyLeuLeuVal 460
DB 405690 GTAGTTGCGCGCGCTTGGAAATCAAGAGAGCTTAAATCAACCAAGTTGCTGCTT 405749
QY 461 IleCysGlySerIleGlyAlaMetThrProTrpLeuAlaValIleLeuPheIleIle 480
DB 405750 ATCTGTGAAGTATGAGCTATGACCTTATCTTGGAGTATCTTCTTTCATCAT 405809
QY 481 AlaIleTrpLeuValSerAlaThrLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeu 500
DB 405810 GCATATTGTTGGTTTCTGCACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 405869
QY 501 LysGluGlnGluValAlaGlnGluAspSerAlaProAlaSerSer 515
DB 405870 AAGAAACAAGAGTGGCTCAAGAAATTAAGTCTCCGCTTCTCA 405914

RESULT 2
US-09-198-452A-1/c
Sequence 1, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Glaxo, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragm
TITLE OF INVENTION: chlamydia and uses thereof, in particular for the diagnosis, f
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A

ALIGNMENTS

C94

162 (e)

RESULT 1
US-09-198-452A-369
Sequence 369, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffois, R.

TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 369
LENGTH: 524
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-369

Query Match 99.5%; Score 2617; DB 4; Length 524;
Best Local Similarity 99.6%; Pred. No. 6,4e-274;
Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 MTKTEKPGKLRSLMPTHTHEKLVPMFLMPCITFNNTVLRDQTLVGPAGSGA 60
DB 10 MTKTEKPGKLRSLMPTHTHEKLVPMFLMPCITFNNTVLRDQTLVGPAGSGA 69
QY 61 EAIPIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPFLFFALFPVITYPLRDVL 120
DB 70 EAIPIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPFLFFALFPVITYPLRDVL 129
QY 121 HPTFADRLQALIPPGILGLVAILRNMTFAAFYVLAELMGVMSLMFPGANEITKHE 180
DB 130 HPTFADRLQALIPPGILGLVAILRNMTFAAFYVLAELMGVMSLMFPGANEITKHE 189
QY 181 AKRYFALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLMAMTIVSGVLMA 240
DB 190 AKRYFALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLMAMTIVSGVLMA 249
QY 241 SYMMINKNVLTDPFYNPEEMQKKGAKPRANKKDSFLYIDRSPYLLLTLLVIAVIGC 300
DB 250 SYMMINKNVLTDPFYNPEEMQKKGAKPRANKKDSFLYIDRSPYLLLTLLVIAVIGC 309
QY 301 INLIEVWKSQKLQYFNNNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKFGMLTGALV 360
DB 310 INLIEVWKSQKLQYFNNNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKFGMLTGALV 369
QY 361 TPTVVLITGIVFPAIVIFRQASGLVAMFGTPTPLMAVVGAILNLSKSTKALDSTK 420
DB 370 TPTVVLITGIVFPAIVIFRQASGLVAMFGTPTPLMAVVGAILNLSKSTKALDSTK 429
QY 421 EMAVYPLDQOKYKGAIDVVAARFGSGGALIOGGLVIGSGISGAMTPYLAIVLLFTI 480
DB 430 EMAVYPLDQOKYKGAIDVVAARFGSGGALIOGGLVIGSGISGAMTPYLAIVLLFTI 489
QY 481 AIVVVSATKLNKLFALQSAIKQOEVAQOESAPASS 515
DB 490 AIVVVSATKLNKLFALQSAIKQOEVAQOESAPASS 524

```

polypeptides, fragments
the diagnosis, prev

TITLE OF INVENTION: and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 653
LENGTH: 551
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-653

Query Match 39.0%; Score 1025.5; DB 4; Length 551;
Best Local Similarity 39.4%; Pred. No. 6e-102;
Matches 214; Conservative 107; Mismatches 185; Indels 37; Gaps 9;

```

QY 1 MTKTEKPGKLRSLMPTHTHEKLVPMFLMPCITFNNTVLRDQTLVGPAGSGA 60
DB 12 MOSSEVKFSLRAYLCPIYKSEFSKVPFLFLAFVGFNYCLKNKMDTLVIGSDAGA 71
QY 61 EAIPIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPFLFFALFPVITYPLRDVL 120
DB 72 EAIPIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPFLFFALFPVITYPLRDVL 131
QY 121 HPTFADRLQALIPPGILGLVAILRNMTFAAFYVLAELMGVMSLMFPGANEITKHE 180
DB 132 HPTFADRLQALIPPGILGLVAILRNMTFAAFYVLAELMGVMSLMFPGANEITKHE 191
QY 181 AKRYFALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLMAMTIVSGVL 238
DB 192 AKRYFALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLMAMTIVSGVL 250
QY 239 MASVWINKNV--LTDPFYNPEEMQKKGAKPRANKKDSFLYIDRSPYLLLTLLVIAVIGC 287
DB 251 MASVWINKNV--LTDPFYNPEEMQKKGAKPRANKKDSFLYIDRSPYLLLTLLVIAVIGC 306
QY 288 LILTLVIAVIGCINTLIEVWKSQKLQYFNNNDYSEFMGNFSFWTGVSVLIMLFVGN 347
DB 307 LILTLVIAVIGCINTLIEVWKSQKLQYFNNNDYSEFMGNFSFWTGVSVLIMLFVGN 366
QY 348 VIRKFGMLTGALVTPMVLITGIVFPAIVIFRQASGLVAMFGTPTPLMAVVGAILNLSK 406
DB 367 VIRKFGMLTGALVTPMVLITGIVFPAIVIFRQASGLVAMFGTPTPLMAVVGAILNLSK 426
QY 407 LSKSTYALFDSSTKENAVYPLDQOKYKGAIDVVAARFGSGGALIOGGLVIGSGISG 466
DB 427 LSKSTYALFDSSTKENAVYPLDQOKYKGAIDVVAARFGSGGALIOGGLVIGSGISG 486
QY 467 AMTPYLAIVLLFTI AIVVVSATKLNKLFALQSAIKQOEVAQOESAPASS 515
DB 487 AMTPYLAIVLLFTI AIVVVSATKLNKLFALQSAIKQOEVAQOESAPASS 546
QY 547 EWA 549
DB 547 EWA 549

```

RESULT 3
US-09-198-452A-306
Sequence 306, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffois, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 306
LENGTH: 907
TYPE: PRT
ORGANISM: Chlamydia pneumoniae

THIS PAGE BLANK (USPTO)

ALIGNMENTS

RESULT 1

ADP, ATP carrier protein CP0408 (imported) - Chlamydia pneumoniae (strain CHL029 at E72089
 C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C/Accession: E72089, D81580
 R.Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999.
 A/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A/Reference number: A12000; MUID:99206506; PMID:10192388
 A/Accession: E72089
 A/Molecule type: DNA
 A/Residues: 1-515 <ARN>
 A/Cross-references: GB:AE001619; GB:AE001363; NID:94376620; PIDN:AA018495.1; PID:9437662
 A/Experimental source: strain CHL029
 R.Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A/Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A/Reference number: A81500; MUID:20150255; PMID:10684935
 A/Accession: D81580
 A/Molecule type: DNA
 A/Residues: 1-515 <REA>
 A/Cross-references: GB:AE002202; GB:AE002161; NID:97189324; PIDN:AA038252.1; PID:9718933
 A/Experimental source: strain AR39, HL cells
 C/Genetics:
 A/Status: adt. 1; CP0408
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 99.5%; Score 2617; DB 2; Length 515;
 Best Local Similarity 99.6%; Pred. No. 8,7e-193;
 Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

102-61
101 gpr

1 MTKTEKPGKLSFMPPIHTEHKKVLPWFMFCITFNVTYLRDRTKTLIVGAPGSA 60
 1 MTKTEKPGKLSFMPPIHTEHKKVLPWFMFCITFNVTYLRDRTKTLIVGAPGSA 60
 61 EAIPFIKFWLVPCALIFMLIVAKLSNLSKQALFYAVGTPFLIFPALPFTVIYPLRDVL 120
 61 EAIPFIKFWLVPCALIFMLIVAKLSNLSKQALFYAVGTPFLIFPALPFTVIYPLRDVL 120
 121 HPTFADRLQALIPGGLGVALIRNMTFAFYVLAELMGSVMSLMFPGFANEITKHE 180
 121 HPTFADRLQALIPGGLGVALIRNMTFAFYVLAELMGSVMSLMFPGFANEITKHE 180
 181 AKRFYALFGIGANISLASGRAIVMAASKLRASVSEGVDPGICSLRLMAMTIVSGVLMA 240
 181 AKRFYALFGIGANISLASGRAIVMAASKLRASVSEGVDPGICSLRLMAMTIVSGVLMA 240
 241 SYMMINKNVLTDPRFYNPEEMQKKGAKPRMKNKDSFLYLDSPYLLTLVIAVYIC 300
 241 SYMMINKNVLTDPRFYNPEEMQKKGAKPRMKNKDSFLYLDSPYLLTLVIAVYIC 300
 301 INLEVTWKSQKIQYPMNDYSEFMGNFSFMTGVSVLIMLVGAVIRKFGMLTGALV 360
 301 INLEVTWKSQKIQYPMNDYSEFMGNFSFMTGVSVLIMLVGAVIRKFGMLTGALV 360
 361 TPWVLLTGIVFPFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTK 420
 361 TPWVLLTGIVFPFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTK 420
 421 EMAVIFLDQEKVKGKAAIDVVAARFGKSGGALLIQGGLVIVCGSIGMTPTLAVITLFTI 480
 421 EMAVIFLDQEKVKGKAAIDVVAARFGKSGGALLIQGGLVIVCGSIGMTPTLAVITLFTI 480
 481 AIWIVSATKLNKFLAQSALKEQVADSDAPASS 515
 481 AIWIVSATKLNKFLAQSALKEQVADSDAPASS 515

RESULT 2

ADP/ATP translocase (imported) - Chlamydia pneumoniae (strain J138) E86534
 C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C/Accession: E86534
 R.Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Nucleic Acids Res. 28, 2311-2314, 2000
 A/Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
 A/Reference number: A86491; MUID:20330349; PMID:10871362
 A/Accession: E86534
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-515 <STO>
 A/Cross-references: GB:BA000008; NID:98978723; PIDN:BA98559.1; GSPDB:GN00142
 A/Experimental source: strain J138
 C/Genetics:
 A/Status: adt. 1
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 99.2%; Score 2610; DB 2; Length 515;
 Best Local Similarity 99.4%; Pred. No. 3e-192;
 Matches 512; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 MTKTEKPGKLSFMPPIHTEHKKVLPWFMFCITFNVTYLRDRTKTLIVGAPGSA 60
 1 MTKTEKPGKLSFMPPIHTEHKKVLPWFMFCITFNVTYLRDRTKTLIVGAPGSA 60
 61 EAIPFIKFWLVPCALIFMLIVAKLSNLSKQALFYAVGTPFLIFPALPFTVIYPLRDVL 120
 61 EAIPFIKFWLVPCALIFMLIVAKLSNLSKQALFYAVGTPFLIFPALPFTVIYPLRDVL 120
 121 HPTFADRLQALIPGGLGVALIRNMTFAFYVLAELMGSVMSLMFPGFANEITKHE 180
 121 HPTFADRLQALIPGGLGVALIRNMTFAFYVLAELMGSVMSLMFPGFANEITKHE 180
 181 AKRFYALFGIGANISLASGRAIVMAASKLRASVSEGVDPGICSLRLMAMTIVSGVLMA 240
 181 AKRFYALFGIGANISLASGRAIVMAASKLRASVSEGVDPGICSLRLMAMTIVSGVLMA 240
 241 SYMMINKNVLTDPRFYNPEEMQKKGAKPRMKNKDSFLYLDSPYLLTLVIAVYIC 300
 241 SYMMINKNVLTDPRFYNPEEMQKKGAKPRMKNKDSFLYLDSPYLLTLVIAVYIC 300
 301 INLEVTWKSQKIQYPMNDYSEFMGNFSFMTGVSVLIMLVGAVIRKFGMLTGALV 360
 301 INLEVTWKSQKIQYPMNDYSEFMGNFSFMTGVSVLIMLVGAVIRKFGMLTGALV 360
 361 TPWVLLTGIVFPFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTK 420
 361 TPWVLLTGIVFPFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTK 420
 421 EMAVIFLDQEKVKGKAAIDVVAARFGKSGGALLIQGGLVIVCGSIGMTPTLAVITLFTI 480
 421 EMAVIFLDQEKVKGKAAIDVVAARFGKSGGALLIQGGLVIVCGSIGMTPTLAVITLFTI 480
 481 AIWIVSATKLNKFLAQSALKEQVADSDAPASS 515
 481 AIWIVSATKLNKFLAQSALKEQVADSDAPASS 515

RESULT 3

probable adp/ATP translocase - Chlamydia trachomatis (serotype D, strain W3/Cx) C71561
 C/Species: Chlamydia trachomatis
 C/Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 20-Aug-1999
 C/Accession: C71561
 R.Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche Science 282, 754-759, 1999
 A/Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
 A/Reference number: A71570; MUID:9900809; PMID:9784136
 A/Accession: C71561
 A/Status: preliminary

THIS PAGE BLANK (USPTO)